

Frontiers in psychodynamic neuroscience

Edited by

Filippo Cieri, Robin Carhart-Harris, Christoph Mathys, Oliver Hugh Turnbull
and Mark Leonard Solms

Published in

Frontiers in Human Neuroscience



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83252-100-7
DOI 10.3389/978-2-83252-100-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Frontiers in psychodynamic neuroscience

Topic editors

Filippo Cieri — Lou Ruvo Center for Brain Health, Neurological Institute, Cleveland Clinic, United States

Robin Carhart-Harris — Imperial College London, United Kingdom

Christoph Mathys — Aarhus University, Denmark

Oliver Hugh Turnbull — Bangor University, United Kingdom

Mark Leonard Solms — University of Cape Town, South Africa

Citation

Cieri, F., Carhart-Harris, R., Mathys, C., Turnbull, O. H., Solms, M. L., eds. (2023).

Frontiers in psychodynamic neuroscience. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83252-100-7

Table of contents

- 05 **Editorial: Frontiers in psychodynamic neuroscience**
Filippo Cieri, Robin Lester Carhart-Harris, Christoph Mathys, Oliver Turnbull and Mark Solms
- 08 **Can Neuroscientists Test a New Physicalist Mind/Body View: DiCoToP (Diachronic Conjunctive Token Physicalism)?**
Linda A. W. Brakel
- 17 **The Affective Neuroscience of Sexuality: Development of a LUST Scale**
Jürgen Fuchshuber, Emanuel Jauk, Michaela Hiebler-Ragger and Human Friedrich Unterrainer
- 27 **Memory for the Future: Psychodynamic Approach to Time and Self Through the Default Network**
Filippo Cieri
- 39 **Psychoanalytic psychotherapies and the free energy principle**
Thomas Rabeyron
- 52 **The nested hierarchy of self and its trauma: In search for a synchronic dynamic and topographical re-organization**
Andrea Scalabrini, Clara Mucci and Georg Northoff
- 73 **Two modes of being together: The levels of intersubjectivity and human relatedness in neuroscience and psychoanalytic thinking**
Riccardo Williams and Cristina Trentini
- 86 **Potential epigenetic mechanisms in psychotherapy: a pilot study on DNA methylation and mentalization change in borderline personality disorder**
Yamil Quevedo, Linda Booij, Luisa Herrera, Cristobal Hernández and Juan Pablo Jiménez
- 98 **An economic model of the drives from Friston's free energy perspective**
Gustaw Sikora
- 114 **Therapist reactions to patient personality: A pilot study of clinicians' emotional and neural responses using three clinical vignettes from in treatment series**
Annalisa Tanzilli, Cristina Trentini, Alessandro Grecucci, Nicola Carone, Chiara Ciacchella, Carlo Lai, Miguel David Sabogal-Rueda and Vittorio Lingiardi
- 134 **Hyperactivity of the default mode network in schizophrenia and free energy: A dialogue between Freudian theory of psychosis and neuroscience**
Jessica Tran The, Jean-Philippe Ansermet, Pierre J. Magistretti and Francois Ansermet

- 152 **Neural correlates of psychodynamic and non-psychodynamic therapies in different clinical populations through fMRI: A meta-analysis and systematic review**
Nicoletta Cera, Jessica Monteiro, Roberto Esposito, Giulia Di Francesco, Dietmar Cordes, Jessica Z. K. Caldwell and Filippo Cieri
- 174 **The Active Inference Model of Coherence Therapy**
D. Eric Chamberlin
- 191 **Differential neural response to psychoanalytic intervention techniques during structural interviewing: A single-case analysis using EEG**
Anna Buchheim, Otto F. Kernberg, Nikolaus Netzer, Peter Buchheim, Corinna Perchtold-Stefan, Barbara Sperner-Unterweger, Fabian Beckenbauer and Karin Labek
- 202 **People solve rebuses unwittingly—Both forward and backward: Empirical evidence for the mental effectiveness of the signifier**
Giulia Olyff and Ariane Bazan
- 222 **From childhood trauma to hyperarousal in adults: The mediating effect of maladaptive shame coping and insomnia**
Frans Schalkwijk, Eus J. W. Van Someren, Nelleke J. Nicolai, Julia L. Uijttewaal and Rick Wassing
- 231 **Converging theories on dreaming: Between Freud, predictive processing, and psychedelic research**
Michael Koslowski, Max-Pelgrom de Haas and Tamara Fischmann
- 240 **Running away from phonological ambiguity, we stumble upon our words: Laboratory induced slips show differences between highly and lowly defensive people**
Lola Thieffry, Giulia Olyff, Lea Pioda, Sandrine Detandt and Ariane Bazan



OPEN ACCESS

EDITED AND REVIEWED BY
Lutz Jäncke,
University of Zurich, Switzerland

*CORRESPONDENCE
Filippo Cieri
✉ cierif3@ccf.org

SPECIALTY SECTION
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 21 February 2023
ACCEPTED 13 March 2023
PUBLISHED 22 March 2023

CITATION
Cieri F, Carhart-Harris RL, Mathys C, Turnbull O
and Solms M (2023) Editorial: Frontiers in
psychodynamic neuroscience.
Front. Hum. Neurosci. 17:1170480.
doi: 10.3389/fnhum.2023.1170480

COPYRIGHT
© 2023 Cieri, Carhart-Harris, Mathys, Turnbull
and Solms. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Frontiers in psychodynamic neuroscience

Filippo Cieri^{1*}, Robin Lester Carhart-Harris^{2,3},
Christoph Mathys^{4,5}, Oliver Turnbull⁶ and Mark Solms⁷

¹Department of Neurology, Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, United States, ²Neuroscape, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States, ³Centre for Psychedelic Research, Department of Brain Sciences, Imperial College London, London, United Kingdom, ⁴Interacting Minds Centre, Aarhus University, Aarhus, Denmark, ⁵Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Zurich, Switzerland, ⁶School of Psychology, College of Human Sciences, Prifysgol Bangor University, Bangor, United Kingdom, ⁷Neuroscience Institute, The University of Cape Town, Cape Town, South Africa

KEYWORDS

psychodynamic, psychodynamic neuroscience, cognitive neuroscience, psychoanalysis, neuropsychanalysis, default network DN, neural networks, Freud

Editorial on the Research Topic Frontiers in psychodynamic neuroscience

Introduction

Parts of cognitive neuroscience still avoid the individual-subjective factor from their investigations, thereby excluding the brain's most unique feature, its selfhood. Many researchers in this field remain skeptical about the psychoanalytic approach and theories. Likewise, many psychoanalysts continue to eschew the structure and functions of the brain in their conceptions of the mind in physiological and pathological conditions. Both cases seem to preserve a Cartesian approach in which the mind is linked to the brain in some arcane manner, detaching subjectivity from the body—rather than treating it an integral part of the complex and dynamic organism as a whole. This approach gives rise on the one hand to a *mindless neuroscience*, and, on the other hand, to a *brainless psychoanalysis* (Cera et al.; Cieri). This Research Topic is an attempt to adopt a psychodynamic approach to cognitive neuroscience, and to furnish a natural science of psychoanalysis (Freud et al., 1954; Solms, 2020).

Chapter 1: Hypotheses and theories

Since the renewed interest in the dialogue between neuroscience and psychoanalysis, especially after the advent of neuropsychanalysis and studies on the Default Network (DN), Cieri proposes a philosophical and scientific proximity between time and self, suggesting a neural overlap in the DN. The author presents studies in cognitive neurosciences and functional Magnetic Resonance Imaging (rs-fMRI) to support his hypothesis. The ontogenetic development of self and time perception is discussed, consistent with the development of the DN's function. Alzheimer's disease is proposed as an example where perception of time is brutally impaired together with a loss of the self's functions.

Rabeyron is interested in a dialogue between psychodynamic therapy and the Free Energy Principle (FEP). Analytically oriented psychotherapy requires a setting, a particular mental state, and specific processes, to induce psychic transformations that can be seen in the light of the FEP. The paper supports a dialogue between psychoanalysis and the neuroscience of subjectivity.

Following a similar idea, **Sikora** is interested in the link between the theory of drives and the Fristonian FEP, with its hierarchical model of the brain, as a helpful integration of the Freudian economic approach. The author uses two psychodynamic traditions: the French and the post-Kleinian school of British psychoanalysis, as one side of the dialogue, with the FEP and the hierarchical predictive model on the other.

In their contribution, **Scalabrini et al.** focus on the nested hierarchy of self and its trauma, suggesting synchrony as a key factor in the processes of dialogue between self and others, which shapes the brain–body–mind system of the subject, including their sense of self. The authors use both Northoff's conceptualization of the brain-based nested three-layer hierarchical structure, and the three levels of trauma theorized by Mucci.

With a more philosophical approach, **Brakel** investigates the mind/body problem through a philosophy of mind framework, introducing issues for dualists and physicalists, along with key concepts such as independent mental causation, emergence, and multiple realization. To respond to some of these problems in a new light, this manuscript proposes a new mind/body approach: the Diachronic Conjunctive Token Physicalism (DiCoToP).

More interested in the psychotherapeutic technique, **Chamberlin** introduces the Active Inference Model of Coherence Therapy, based on Psychological Constructivism. This form of therapy suggests Coherence Therapy as a dyadic act of therapist guided Active Inference which renders conscious the potential unconscious causes of a patient's behavior. The author proposes this approach as a computational process useful for therapeutic help and experimental research design.

Chapter 2: Original research

In their work, **Fuchshuber et al.** are interested in developing a standardized self-rated evaluation for the LUST emotion. The authors produce two versions of the L-scales (L-12 and L-5). Cronbach's α indicates good internal consistency regarding L-12 ($\alpha = 0.90$), and acceptable value for the L-5 ($\alpha = 0.82$). These scales demonstrate satisfying psychometric properties.

Quevedo et al. propose an experimental approach to explore a different field: the gene-environment interaction in the development of Borderline Personality Disorder (BPD). The authors use an epigenetic approach to show how molecular machinery adapts to the environment. They use a pilot study, with a small sample of BPD patients, exploring changes in peripheral DNA methylation of the FKBP5 gene, which encodes for a stress response protein, in relation to psychotherapy (both symptomatology and underlying psychological processes).

Bazan et al. have contributed to this Research Topic with two separate works. In the first, **Olyff and Bazan** recruited 1,458 participants using a rebus priming paradigm, where the images were followed by a target word semantically related to the rebus

resolution, upon which the participants, unaware of the rebus principle, produced six written associations. The authors show how the images induced inadvertent rebus priming in naïve participants, suggesting that people solve rebuses unwittingly and independently of stimulus order, thereby constituting empirical evidence for the mental effectiveness of the signifier.

In their second experimental contribution, **Thieffry et al.** underline how the defenses measured in their research are internal, intimate control systems, probing for the censorship between the systems Unconscious and Preconscious. This study contributes support for a psychodynamic explanatory model of the production of parapraxes.

Schalkwijk et al. focus their research on Adverse Childhood Experiences (ACE's), potentially indexing vulnerability to maladaptive coping and stress, associated with insomnia. The authors used existing data from subjects with insomnia and normal controls, asking participants to complete the questionnaire about traumatic experiences during childhood, to explore the association between ACE's, shame coping-styles, adult insomnia, hyperarousal, and the neurobiology of autobiographical memory. Their findings can have implications for the treatment of insomnia, with more focus on traumatic experience and emotional processing rather than the typical (more superficial) sleep interventions.

Tanzilli et al. indicate that therapists highlighted different patterns of criticized/devalued and sexualized reactions to visual images of patients with distinct personality disorders, at statistically systematic and clinically meaningful levels. Moreover, psychotherapists' late positive potentials (LPPs) in the hippocampus were able to determine which patient they observed during the EEG task, with high accuracy.

Still in the field of the psychoanalytic setting and the investigation of brain correlates, **Buchheim et al.** explored change in the electroencephalographic (EEG) signal as an effect of psychoanalytical therapeutic interventions, investigating brain correlates of specific psychodynamic approaches in the EEG power spectrum. The authors contrast three types of intervention (clarification, confrontation, and interpretation) and a neutral control condition during a structured psychoanalytic interview conducted while the EEG was recorded. The authors were able to show that incisive interventions, such as confrontation with discrepancies and interpretation of unconscious intrapsychic conflicts, can stimulate temporary emotional lability, causing a change in psychic processing akin to interference from external stimuli.

Chapter 3: Review, systematic review, and meta-analysis

Cera et al. use the approach of systematic review and meta-analysis to examine neural changes after psychotherapy, in several different mental disorders. The authors also attempt an exploration of the different psychotherapies' approaches, particularly interested in the comparison between psychodynamic and non-psychodynamic approaches. The meta-analysis and systematic review found that all psychological interventions influence the brain from a functional point of view, showing their effects from a neurological perspective. Frontal and prefrontal

regions, insular cortex, superior and inferior frontal gyrus, and putamen are all involved in these changes following psychological therapy. Psychodynamic approaches are more prone to evoke changes in the latter three regions.

Williams and Trentini present an overview of the main contributions on intersubjectivity in the field of neuroscience. Based on the capacity for emotional resonance, a primary sense of connectedness is proposed by the authors, and potentially defined as intersubjective in that it entails shared affective states and intentions with caregivers. The authors propose to think of this kind of intersubjectivity as contingent. They also propose a multi-layered approach to intersubjectivity, consistent with current neuroscientific conceptualizations.

Tran The et al., are interested in the DN and how it can offer a useful field of investigation in the dialogue between mind and brain, particularly in the field of schizophrenia. Combining neuroimaging studies with Freudian hypotheses, the authors propose that hyperactivity of the DN is a consequence of a process of massive reassociation of traces in schizophrenia. This process may constitute an attempt at minimizing the excessive free energy that is present in psychosis.

Koslowski et al. try to connect the field of dream research, the predictive processing account of human cognition, affective neuroscience, neuropsychanalysis, and emerging research on psychedelic substances, to deepen our understanding of the mechanisms of the dreaming brain and dream-like states. Conceptual bridges between theories of consciousness, dream research and neurobiological accounts are proposed, to further advance empirical studies on the nature of, and different functions of, dreaming.

Conclusions

As editors, we are impressed by the quantity and quality of these contributions, often with an empirical—rather than

only a speculative—approach to the problems proposed by this Research Topic. The psychodynamic approach in neuroscience—and dialogue between neuroscience and psychoanalysis more generally—is complex, and it holds a lot of open and challenging questions for future research. To disentangle this complexity, or at least to address it, the combined effort of multiple scientific approaches is needed—as underlined in this Research Topic. New approaches in formal education would also be useful, to build up a new generation of researchers, clinicians, and scientists, equipped with notions about the neuroscientific approach to psychoanalysis, and about the psychodynamic approach to cognitive neuroscience, recognizing—as mentioned—that subjectivity is an integral part of the complex and dynamic organism as a whole.

Author contributions

FC: wrote the editorial. RC-H, CM, OT, and MS: read, corrected, commented, and approved the paper. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Freud, S., Bonaparte, M., Freud, A., and Kris, E. (1954). "Project for a scientific psychology," in *The Origins of Psycho-analysis: Letters to Wilhelm Fliess, Drafts and Notes: 1887-1902* (Transl. by E. Mosbacher, J. Strachey; New York City: Basic Books/Hachette Book Group), 347–445. doi: 10.1037/11538-013
- Solms, M. (2020). New project for a scientific psychology: general scheme. *Neuropsychanalysis* 22, 5–35. doi: 10.1080/15294145.2020.1833361



Can Neuroscientists Test a New Physicalist Mind/Body View: DiCoToP (Diachronic Conjunctive Token Physicalism)?

Linda A. W. Brakel^{1,2,3*}

¹ Department of Philosophy, University of Michigan, Ann Arbor, MI, United States, ² Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States, ³ Michigan Psychoanalytic Institute, Farmington Hills, MI, United States

Given that disparate mind/body views have interfered with interdisciplinary research in psychoanalysis and neuroscience, the mind/body problem itself is explored here. Adding a philosophy of mind framework, problems for both dualists and physicalists are presented, along with essential concepts including: independent mental causation, emergence, and multiple realization. To address some of these issues in a new light, this article advances an original mind/body account—Diachronic Conjunctive Token Physicalism (DiCoToP). Next, puzzles DiCoToP reveals, psychoanalytic problems it solves, and some empirical evidence accrued for views consistent with DiCoToP are presented. In closing, this piece challenges/appeals for neuroscience research to gain evidence for (or against) the DiCoToP view.

Keywords: mind/body problem, token physicalism, emergence, multiple realization, neuronal assembly, mental causation, Diachronic Conjunctive Token Physicalism

OPEN ACCESS

Edited by:

Filippo Cieri,
Cleveland Clinic Lou Ruvo Center
for Brain Health, United States

Reviewed by:

Enrico Facco,
University of Padua, Italy
George F. R. Ellis,
University of Cape Town, South Africa

*Correspondence:

Linda A. W. Brakel
brakel@med.umich.edu

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 29 September 2021

Accepted: 24 November 2021

Published: 17 December 2021

Citation:

Brakel LAW (2021) Can
Neuroscientists Test a New
Physicalist Mind/Body View: DiCoToP
(Diachronic Conjunctive Token
Physicalism)?

Front. Hum. Neurosci. 15:786133.
doi: 10.3389/fnhum.2021.786133

INTRODUCTION

Frontiers Research Topic, Psychodynamic Neuroscience, calls for a deepening dialogue between neuroscience and psychoanalysis. This is exciting, just as it stands. However, one further participant discipline, the philosophy of mind, could offer much toward addressing the conflicting mind/body views inhibiting progress both in psychoanalysis and neuroscience. It is within this added discipline that I attempt a contribution to this collection. The article consists of four sections, and then a final discussion, with conclusions and summary.

Section One explores essential problems embedded within the mind/body problem, arriving in Section Two at the view I devised (Brakel, 2013) and now advocate—Diachronic Conjunctive Token Physicalism (DiCoToP)¹. Although the DiCoToP account resolves some of the issues associated with the mind/body problem, Section Three grapples with new puzzles that this particular mind/body account reveals and entails. Section Four, the final section, has two parts. In Part One, I aver that DiCoToP can go some distance in explaining two matters endemic to clinical psychoanalytic work but heretofore quite vexing. These are: (a) working through; and (b) the mechanism of therapeutic action after patients gain new knowledge through interpretation. Part Two, acknowledges that even if I am correct about these and more general aspects of DiCoToP's explanatory power, much more

¹ Note, I have taken up the mind/body issue in much greater length and depth in Brakel (2013). Much will be abbreviated and condensed here.

than such coarse correlations are needed. Hence, after providing a brief account of some existing empirical studies whose findings can be taken to clearly support DiCoTop, the close of Part Two is essentially an appeal (and a challenge) to neuroscientists. Can neuroscientists apply cutting edge neuroscientific methods to explore in humans whether or not features of DiCoToP could obtain? Whatever the outcome, I would hope that this sort of empirical investigation would be a thoroughly interdisciplinary adventure, totally worth taking, as is the current research topic itself.

SECTION ONE

The Mind/Body Puzzle: Problems for Physicalists, and Dualists

“Why physicalism?” is the first question every physicalist² must consider—this, especially when dualism seems so appealing. Dualism’s appeal owes to its rather elegant solution to some basic and difficult mind/body problems for physicalists (especially physicalists of the reductive variety)—that of independent mental downward causation. Thus, as reductive physicalists maintain, (a) there is nothing over and above the physical (and the laws thereof), and (b) everything mental is at base physical; how then can mental goings-on—familiar processes of the mind, like qualia, beliefs, desires—be genuine causes of behavior, rather than just epiphenomena? Relatedly, dualism allows a clear, unproblematic grasp of such important concepts and phenomena as the Self, subjective first personal accounts, and even consciousness itself.

And yet, for physicalists, dualism presents its own insurmountable problem: If one embraces the most prevalent scientific view of a physical-only universe, the very nature of the mental—ontologically, metaphysically, as well as epistemologically—has not been, and perhaps cannot be, satisfactorily accounted for. Nonetheless, and notwithstanding this problem, there are two famous and compelling arguments for dualism. I will discuss them both in the section below, along with my claim that they do not work!

The Dualists Have Two Great Arguments; But They Don’t Work The Knowledge Argument

The Knowledge Argument was advanced by Frank Jackson in 1982. The main character is Mary, the color scientist. Jackson (1982, p.130) describes Mary as “...a brilliant scientist who is...forced to investigate the world from a black and white room via a black and white television monitor. She specializes in the neurophysiology of [color] vision and acquires...all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like ‘red,’ ‘blue,’ and so on.” Jackson (p.130) then asks two pivotal questions: What will happen when Mary is released from her room and sees colors for the first time in her life? “Will she learn anything or not?” He concludes (p.130): “It seems just obvious that she will

learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.”

This argument, although clever, compelling, and even convincing to many, is (on my view) fatally flawed. Indeed, one can assume Mary did not know what it’s like to experience color. But then comes the problematic part. For the argument to go through, one must also assume that knowing-what-it-is like type knowledge—so called acquaintance knowledge (or know-how knowledge)—is not part of physical knowledge. In other words, the argument assumes at the front end, exactly what it is trying to prove: namely that there exists knowledge that is non-physical.

The Zombie/Conceivability Argument

The Zombie/Conceivability Argument is presented concisely by Chalmers (2010, p.106). Here he paraphrases his earlier work (and that of others), setting up the Zombie Case as follows: “...It is conceivable that there be a system that is physically identical to a conscious being but lacks...that being’s conscious states. Such a system might be a zombie: a system that is physically identical to a conscious being but that lacks consciousness entirely.” He continues (p.107), that such a being/system

...will look identical to a normal conscious being from the third-person perspective. In particular...brain processes will be molecule-for-molecule identical with the original, and their behavior will be indistinguishable. But things will be different from the first-person point of view. What it is like to be...a...zombie will differ from what it is like to be the original being. [For] there is nothing it is like to be a zombie.

The next step in the Zombie Case, according to Chalmers (p. 107), is to acknowledge that since such zombie systems/beings are coherently imaginable with “no contradiction,” it can be inferred that they are a metaphysical possibility: “From here, it is inferred that consciousness must be non-physical. If there is a metaphysically possible universe that is physically identical to ours but that lacks consciousness, then consciousness must be a further, non-physical component of our universe.” Again, quoting Chalmers (p.107):

- (1) It is conceivable that there are zombies.
- (2) If it is conceivable that there are zombies, it is metaphysically possible that there are zombies.
- (3) If it is metaphysically possible that there are zombies, then consciousness is non-physical.
- (4) Consciousness is non-physical.

This argument for dualism too, is very appealing. And yet, there are serious problems embedded within its structure. Several authors, including Jackson (2003, pp.9, 30) and (Nagel, 1998, p.346), suggest that the Zombie Argument, turning on the conceptual conceivability of conscious-less human physical duplicates, is paradoxically a function of our own human concepts, and the limitations therewith (See also Hill, 1997). Thus, our powers of conceiving may outstrip both metaphysical reality and epistemological possibility. A physicalist can then mount a simple 6 step counter argument.

²Physicalism is synonymous with Materialism in the mind/body context; as such they will be used interchangeably here.

- (1) Person-X and Zombie-X are molecule-for-molecule/neuron-for-neuron identical;
- (2) Person-X's consciousness is no more than a physical arrangement of these molecules/neurons;
- (3) Zombie-X has the same arrangement of these same molecules/neurons;
- (4) Hence, if Zombie-X does not have consciousness, this must be due to some consciousness-blocker.
- (5) Any blocker must itself be physical, subject to physical laws (since there is nothing over-and-above the physical)³.
- (6) The upshot: Either Zombie-X is not a zombie, and is conscious; or Zombie-X is not conscious, but not identical with Person-X.

With the hope that these refutations of two great dualist arguments have had some sway, let's move on to some physicalist views.

Non-reductive Physicalism: Appealing but Problematic

Non-reductive physicalism has promise. It is a physicalist/materialist view, and yet holds (as its name signals) that the mental is not reducible to the physical. It follows that non-reductive physicalism holds many of the advantages of dualism, while avoiding dualism's biggest challenge—the ontologic nature of the mind and its mental goings-on. For non-reductive physicalists, the mind is strictly material/physical ontologically; but mental properties, e.g., qualia, the experience of sensations, the content of beliefs, consciousness itself, can neither be reduced to, nor explained by physical processes, not even physical laws. On the plus side, as with dualism, non-reductive physicalism allows such mental entities as Self, first person subjectivity, qualia, conscious attitudes and their contents (like beliefs and desires) to have real and independent downward causative powers.

However, there is a sizeable minus side too. If a physicalist maintains that there is nothing (no substance, or process, etc.) beyond or over-and-above the physical, what does the non-reductive solution really amount to? How is non-reductive physicalism different from property dualism? Next, and more important than the classificatory issue is the question of how these mental goings-on—irreducible-to-the physical—actually effect their novel, independent (non-epiphenomenal) causative powers? What laws of nature apply?⁴

³Blockers, described by Leuenberger (2008), are necessary if one is actually conceiving of initial physical identity between Person-X and Zombie-X; this, whether or not one is aware of conceiving of blockers! But Chalmers (2010, p.164) tries to block blockers, “[W]hen we conceive of the zombie world... we conceive of a world with physics and *nothing else* [Chalmers's italics]; no blockers.” Hence Chalmers not only argues that consciousness is non-physical, he extends this to blockers. We must then ask, why would blockers be outside the world of physics?

⁴Similar problems, and more, arise for dual-aspect monists, particularly those who suggest that the monism endorsed is not strictly physical, but algorithmic or symbolic. Also, there are panpsychists, who also hold that there is nothing beyond the physical, but turn things upside down positing that it is the experiential/mental that is clearly and incontrovertibly physical. Further, that this is the case fundamentally, perhaps right down to the level of atomic (and subatomic) particles (See especially Strawson, 1994, 2009). Difficulties here arise in (a) accounting for whatever is non-conscious/non-experiential, and (b) explaining

Emergentism might help the non-reductive physicalist. In simplest terms, emergentism holds that “macroproperties” arising from a number of “microproperties” conjoined, can have autonomous causal effects. For example, many H₂O molecules combined constitute a macroproperty, emerging from a combination of many single H₂O molecules—the microproperties. Note that the macroproperty, water, has properties, different and novel, from the underlying microproperties—the single H₂O molecules. Here the macroproperty, water, is a fluid, a liquid, it is wet. These are emergent properties, and water is causally autonomous and independent from those of the underlying single H₂O molecule microproperties.

Applying this to the mind/body puzzle—emergentists aver that mental macroproperties emerge from physical microproperties. Searle (1992, p.112) states:

...consciousness is a causally emergent property of systems. It is an emergent feature of certain systems in the same way that solidity and liquidity are emergent features of systems of molecules. The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them.

This kind of sanctioned emergence Searle calls “emergent1” and he (p.112) makes a sharp distinction between such allowable emergent1 phenomena and another sort of emergence which Searle abjures. Here is Searle (p.112) on “... a much more adventurous conception, call it ‘emergent2.’ A feature F is emergent2 iff [if and only if] F is emergent1 and has causal powers that cannot be explained by the causal interactions a, b, c...” Searle continues (p.112), “If consciousness were emergent2, then consciousness could cause things that could not be explained by the causal behavior of the neurons.” This would be an impossible outcome that Searle rejects⁵.

But Searle's important discussion of emergentism has not actually solved the problem. Yes, Searle acknowledges that the existence of consciousness is reductively accounted for physically, but to repeat the last clause of his last quoted sentence just above: “...but consciousness cannot itself be deduced [reduced to] or calculated from the sheer physical structure of the neurons without some additional account of the relations between them.” This is true, but Searle does not answer two questions that immediately follow: What are these accounts? And more broadly, can consciousness indeed be deduced or calculated from additional accounts of the relation of neurons, even in principle? If yes, then reductive, rather than non-reductive, physicalism is rescued; but is its cost too high with respect to autonomous mental causal powers? On the other hand, if consciousness

the seeming lack of conscious experience until the fundamental particles undergo much combinatory work. But an examination of these issues is beyond the scope of this article.

⁵Continuing with Searle (p.112) on the very impossibility of emergent2 phenomena, he states: “...it seems unlikely that we will be able to find any features that are emergent2, because the existence of any such features would seem to violate even the weakest principle of the transitivity of causation.” Strawson (2006, pp. 69–71) also dismisses emergent2 emergence, labeling it “radical emergence.”

cannot be so calculated, the notion of independent mental causation seems vindicated; but do these autonomous mental powers arise in the forbidden emergent² fashion?

In the next section, I will present my own view, which features mental goings-on as emergent¹ phenomena, along with the sort of independent causal powers seen with other emergent¹ macroproperties, such as the familiar example of water's liquidity and H₂O molecules?

SECTION TWO

Diachronic Conjunctive Token

Physicalism: DiCoToP

Why Token Physicalism, Not Type Physicalism?

In continuing to explore reductive physicalism, there is first another terminological issue: Reductive physicalism/materialism can also be called identity theory. Next, there is a substantive categorization matter: There are two sorts of reductive physicalists—type and token physicalists. J.C. Smart (1959, p.156), advocates for reductive physicalist “brain process theory [both type and token]” on the basis of “the principles of simplicity and parsimony.” Hill (1991, p.6), a type materialist himself, holds that both type and token physicalists “. . . maintain that we obtain a simpler and more straightforward picture of the universe if we assume that sensory [mental] events are identical with physical events.” Elaborating this further, Hill (p.22) discusses reductive physicalists as having the scientific advantage of “the best explanation principle” whereby, “If a theory provides a good explanation of a set of facts, and the explanation is [as good as or] better than any provided by a competing theory, there is good and sufficient reason for believing that the theory is true.”

Now as to type vs. token physicalism, Hill (p.11) explains: “*Type materialism* is like token materialism in claiming that sensory [mental] events are identical with physical events. However, [different from token materialism], it also claims that there is a set of physical characteristics with which qualitative [mental] characteristics are universally and lawfully correlated. . . .” Continuing, Hill (p.35) goes on to assert that “. . . type materialism is simpler than dualism because it postulates fewer *events*. . . [and] type materialism is simpler than double aspect theory because it postulates fewer *facts* [Hill's italics throughout].” And as far as type materialism over token becoming the more popular view, it is understandable because type materialism brought a universal, lawful regularity to the vexing mind/body problem: predictable mappings of mental state types of a certain character onto brain states types of a certain character. What could go wrong? The answer is multiple realization, the topic of the next subsection.

Multiple Realization

First articulated in a seminal article by Putnam (1967/1975), the Multiple Realization Argument (MR), and its consequences for type physicalism is described in some detail by Kim (1993). According to Kim (p.179) The MR argument advances the idea that given that “. . . any mental state. . . can be ‘physically realized’ in many diverse types of organisms and physical

structures (e.g., humans, mollusks, crustaceans, and perhaps Martians and robots) so that, as a matter of empirical fact, it is extremely unlikely that some uniform physical state exists to serve as. . . physical correlate [to that mental state].” What follows then, according to Kim (p.272):

. . . [Since] any psychological event-type can be “physically realized” or “instantiated” or “implemented” in endlessly diverse ways, depending on the physical-biological nature of the organism or system. . . it [is] highly implausible to expect the event[-type] to correlate uniformly with, and thus be identical with [or reducible to] some “single” type of neural or physical state.

This conclusion from the MR argument made a deep impact upon those working on the mind/body question. Kim (1993) characterizes its effect as profound and devastating to type materialism (p.309): “[T]ype materialism’ is standardly thought to have been definitively dispatched by MR to the heap of obsolete theories of mind.” Indeed, because of this argument, type materialists seem to have retrenched. Certainly they limited their focus, as talk of type materialism narrowed to type materialism within a species. Thus there would be human type pain (Mental-H-P) and human type brain circuits (Physical-H-BC); these not necessarily similar to mollusk type pain (Mental-Mol-P) and mollusk nervous tissue (Physical-Mol-NT); and Martian type pain (Mental-Mar-P) and Martian type physical type pain realizer (Physical-Mar-PR); all of these also not necessarily similar to one another.

But, as I remarked in 2013 (Brakel, pp.75–76) and paraphrase here: MR is problematic even for species-specific type physicalists. Neuroscientists have long established that among different individual humans there are various brain circuits involved in the same mental property—sometimes even gross differences in brain areas. Along with these micro and even macro brain differences across individuals, all realizing the same mental property/event, there are variations within a single individual—this even more problematic for type physicalists. According to Bechtel and Mundale (1999, p.176) within any one person “. . . the same psychological state can be realized by different brain states. . . a many-to-one mapping from brain states to psychological states.” The term for this is “biological degeneracy,” which Edelman and Gally (2001, p.13763) define as “. . . the ability of elements that are structurally different to perform the same function or yield the same [contentful] output.” Noppeney et al. (2004, pp.440–441) report that sometimes such within-individual degenerate plastic brain changes can be dramatic in preserving important psychological functions. Getting even more fine-grained and specific, Figdor (2010) explains (p.435): “On the neuroanatomical side, single cells, neuronal populations, anatomical areas, or anatomical networks are among the ‘structural elements’ that may appear in degenerate mappings.” (See also Price and Friston, 2002; Friston and Price, 2003; Noppeney et al., 2004).

Undoubtedly the brain's degeneracy and its plasticity are of great import for survival and for evolutionary success. Yet, these current scientific developments severely compromise any type physicalist view of the mind/body question. It becomes

increasingly apparent that even the more limited local species-specific type physicalist cannot rely on type-type identity or type-type reduction even within the brain of a single individual, much less across the brains of an entire species. No one-to-one physical type to mental type correlation can be found to exist.

The usual reaction among philosophers to this daunting challenge to type materialism has been to turn back to a non-reductive version of physicalism, neglecting token materialism, which in my view deserves much further investigation. Indeed, in the next subsection another issue will be presented, which while seemingly problematic, also admits of possible solutions offering potential advantages for token physicalism.

The Question of Scale

An understanding of multiple realization leads to another potential problem for physicalist-identity mind/body solutions, token as well as type. That is the issue of scale. There is an obvious mismatch between a mental process/event—e.g., a particular belief—and the myriad of neurons that underlie such a belief, even in a single person at a single time. The brain processes are much more fine-grained. Now add to this the fact that the same specific belief can be realized on a number of different occasions, each with its own neuronal assembly. Some such assemblies are found within degenerate mappings, as described above (See also Friston, 1997; Price and Friston, 1997). But even when neuronal assemblies occur within a localized neural circuit, there is still a one-to-many mismatch. Neuroscientists Papadimitriou et al. (2020, p.14464) characterize neuronal assemblies in general as “...large populations of neurons believed to imprint [particular] memories, concepts, words, and other cognitive information.” They further describe these as “randomly connected,” demonstrating plasticity.

Given this mismatch, how can token physicalism—which essentially holds that one particular mental event is identical with a physical event—hold up? Stephen Yablo (1992, p.256, 271) proposes a view of the relation between mental and physical properties and events, which can be helpfully applied to token physicalism. Yablo (1992, p.256) briefly outlines his account of determinable/determinate relations as follows: “Necessarily, something has a mental property iff [if and only if] it has also a physical determination of that property.” This, he explains, “...is an instance of the standard equation for determinable and determinates, generally, namely, that something has a determinable property if it has some determinate falling thereunder.” The color blue for example is a determinable with various shades of blue—navy blue, oxford blue, pale blue, gray blue, sky blue, turquoise blue, midnight blue, royal blue—being among its determinates. Applying this to the question at hand, a particular belief (and its content) would be the determinable, while several different neuronal assemblies would be its various determinates.

Composites and Conjunctions

The idea of a single determinable with multiple determinates raises another interesting matter. Is a determinable a composite of all of its determinates? Is it a conjunction? Is it different for various determinable/determinate combinations? The question is hard to answer for a determinable like blue—but blue seems to be

more of a mixture of the different blues; a composite, rather than a conjunction. The question is more easily settled for another sort of determinable, The New York Yankees. Since its inception as the team called “The New York Yankees” in 1913,⁶ until its current instantiation in 2021, the New York Yankees are a conjunction (not a composite) of all of the New York Yankee teams during those years. On the other hand, all of the players playing for a specific Yankee team, e.g., 1960, comprise a composite—the 1960 New York Yankees. But, interestingly if one, two, or even all the players had been traded, the 1960 New York Yankees would still have been the 1960 New York Yankees, still a composite, but a different one, consisting of different parts, and different tokens.

Of course, the question here is how do these matters apply to the mind/body problem and token physicalism. I will attempt to answer this in the next subsection.

Diachronic Conjunctive Token Physicalism (DiCoToP)

I have posited (Brakel, 2013, Chapter 3) a novel token physicalist view of the mind/body problem—Diachronic Conjunctive Token Physicalism (DiCoToP). It features a brain-based view of mental content in which each singular mental event (i.e., a particular belief with a specific mental content⁷) occurring at time-*t*, exists as a brain event at time-*t*, consisting of an assembly of neurons, along with whatever neurochemical processes facilitate their connection. This is a reductive token physicalist view, which as such means that at time-*t*, synchronically, there is nothing over and above these brain goings-on as far as the specific mental event is concerned. Further, each additional singular instance of the same mental content (each token)—take for example, “My dog is fun”—is populated by a different network of neurons, perhaps even a different neuronal network, because it occurs at a different time, in a different place, and within a different context. Sometimes the neural assemblies at times-*t* and *t* + 1 or *t* + 20 vary only slightly; sometimes they are quite different. In any case, the sum of all of these instances of this content, i.e., all of their variable neuronal assemblies—neurons, networks, circuits—the conjunction, over time, diachronically, constitutes the mental content.

On this account, the mental event/process would be the determinable and its various neuronal ensemble instantiations would be the determinates. Following Yablo (1992, p.259), this understanding admits of at least a partial resolution of the mental causation problem: “...determinates and their determinable. . .are not causal rivals. . .” And Yablo adds (p.272) “...rather than competing for causal honors, determinable and their determinates seem likelier to share in one another’s successes.”

⁶This team played under different names from 1901–1903 and 1903–1913.

⁷I chose mental content for ease of explanation and potential ease of research investigation. A reviewer made the incontrovertible point that emotions have much to do with shaping cognition, and, that actions are integrally involved with the workings of the brain. Since I totally agree, additional (or other) empirical research would be welcome. For example, perhaps neuronal assemblies responding to basic positive vs. negative stimuli (pain vs. pleasure) would admit of neuroscientific investigation with equal (or perhaps greater) facility than the neutral content I put forward. Actions too—both spontaneous and directed—could be good sources for neuroscientific studies of plastic neuronal assemblies.

But even if these determinables—the mental goings-on—do have autonomous causal powers, there remains the still vexing question of the relation of the physical to the mental, namely how is it that mental goings-on, like consciousness, arise from its physical-only base?⁸

Re-exploring emergence (particularly the sanctioned emergent1 process) now with the DiCoToP model in hand, a solution (at least a partial one) to this overarching puzzle might be achievable. But, what is more certain, is that emergent1 processes viewed in terms of the DiCoToP mind/body model, can add substantial evidence for autonomous mental causation. Recall that macroproperties consisting of collections of microproperties do have real and novel causal powers. A big collection of H₂O molecules has new properties—it is a fluid, it is wet—and with these properties has causal powers independent from single H₂O molecules. Diachronic Conjunctive Token Physicalism maintains that a specific singular mental content repeated over time—e.g., the belief “my dog is fun”—should be considered a macroproperty consisting of the conjunction of the many physical neuronal assembly microproperties entrained over the numerous instances of this particular mental content. Indeed, this fits with Searle’s (1992) suggestion, described above, that Ms are emergent1 macroproperties, micro-based upon Ps; and their relation can therefore, in principle, be explained in the usual way by normal science.

But, to be fair, this account of Ms and Ps is not a universally endorsed. Kim (1998, p.98), for one, does not consider an M to be a macroproperty micro-based on Ps; and even Searle (1992) contests the emergent1 solution as explanatory for first person ontology. That acknowledged, let’s consider what is different in the relation of Ps to Ms, on the one hand, and H₂O molecules to the wet fluid we know as water, on the other. The difference, I aver, is that physical chemists know how the multiple molecules of H₂O, when accumulated give rise to wet water. They can understand the organization of the many molecules within a structure, the Brownian movements of the molecules under temperature and pressure conditions, etc., On the contrary, even if DiCoToP proves to be correct, we would still have no good understanding of the “how” relation between mental contents as a macroproperty and the myriad neuronal assemblies that are its microproperties. Hence, the final subsection of this work will essentially be a mixture of plea and challenge to neuroscience colleagues to work on the “how” within the DiCoToP account. But first, in the section just to follow are some new puzzles the DiCoToP view brings to light.

SECTION THREE

New Problems Revealed/Entailed by Diachronic Conjunctive Token Physicalism

With Diachronic Conjunctive Token Physicalism in place, new puzzles come to the fore. To demonstrate, let’s start with a thought experiment. Suppose that water’s liquidity/wetness which seemingly owes to a very large collection of H₂O

molecules “really” owes to the perceptual capacity of the subject experiencing the wetness. In other words, suppose that for a very different sort of experiencer (one that is microscopically small) only two or three H₂O molecules would suffice as a sufficient microproperty base for the emergent1 macroproperty, water, to arise—this, with its liquidity, wetness, and most importantly, its independent causal powers. The parallel question immediately presents itself: Would a small number of neural assemblies, even just two, allow this imagined tiny experiencing subject to experience consciousness, and allow mental independent (not epiphenomenal) causal efficacy?

Related to this is the Sorites Paradox, a paradox demonstrating the inherent vagueness (lack of specification of the extension) of certain concepts. Here are two simple examples involving the concepts “baldness” and “the heap.” If someone is bald and gets 1 more hair, is he still bald? What about 2? What about increasing by 1 hair iteratively? Likewise, does 1 grain of sand constitute a heap? No, what about 2? What about increasing 1 grain-at-a-time for n-times? Which specific hair makes the difference between baldness and no longer bald? Which grain is the one that constitutes the heap? No one of them can be that potent, and yet at some point there is a heap⁹. Now let’s apply this first to the uncontested macroproperty, water. How many H₂O molecules are necessary for us to perceive H₂O as wet, liquid water? Next, how many neurons and neuronal assemblies are necessary for consciousness to be recognized as such? In addition to the “recognized-as-such-by-whom” problem outlined just above, we now have introduced the notion that the very concept of consciousness might be vague, unspecifiable^{10,11}.

Both of these problems might lead one to the panpsychism family of materialist mind/body views, in which (a) physicalism holds; (b) conscious experience is the only thing we really know for sure; and (c) even the most basic physical particles (including their subatomic constituents) can be mental/conscious (Again, see especially Strawson, 1994, 2009). But as I’ve outlined earlier (footnote 4), these views also have major problems. Here are two: (1) Either there must be an accounting for whatever aspects are non-conscious/non-experiential, if any such parts are posited; or (2) If there is nothing that is non-conscious, the panpsychist needs to explain why it is that our human conscious experience is not experienced until these fundamental “conscious” particles undergo much combinatory work. Further, and of no less importance, the nature of these combinatory processes ought to be outlined.

Paradoxically, it is these somewhat esoteric philosophical puzzles revealed by DiCoToP that can return us to something potentially more solvable—although still very difficult. Here then is the central question:

⁹For more on vagueness and the Sorites Paradox, including basic references, see Brakel (2010), Chapter 3, pp.53–87.

¹⁰So, is the problem of consciousness only a problem for creatures like us, with problematic vague concepts, and the constraints (size and otherwise) of our particular perceptual capacities? In other words, is this a mind-dependent subjective problem writ large? But note, we are creatures like us—we are us—and of course we want to solve the problem for us!

¹¹On the other hand, as suggested by a reviewer, consciousness may be a threshold phenomenon, and as such not so susceptible to the sorites problem.

Given (a) that many H₂O molecules are needed for the emergent¹ property of liquidity and the subjective experience thereof; and (b) that the “how” of liquidity/fluidity can be understood in terms of the principles of physics and physical chemistry.

Q: Can a similar understanding about how the myriad neural assemblies align/combine in emergent 1 fashion to yield consciousness?

So far no, but eventually, with more advances in neuroscience, couldn't the answer evolve: first to “not yet” and later to “yes!”? Meanwhile, the Diachronic Conjunctive Token Physicalist view does now allow some explanation for some heretofore perplexing aspects of psychoanalysis. It is to these I next turn.

SECTION FOUR

Part One: DiCoToP Can Explain A Problem Endemic to Clinical Psychoanalyses

“Working-through” is a concept familiar to psychoanalysts. It represents an important process present in almost every psychoanalytic treatment. Working-through is especially significant theoretically as it can be seen as a shorthand version of perhaps one of the least understood problems for clinical psychoanalyses—namely, what is the mechanism of therapeutic action after interpretative work facilitates new knowledge? To the extent this is the case, comprehending working-through will aid in understanding this second more extensive problem. But even the “simpler” mechanism of working-through, specifically how it actually takes place, has not been well grasped. Thought to be related to grieving, the basic notion is that reviewing/re-working painful, problematic or neurotic contents from different angles helps to effect actual change. The mind/body account I've advanced here and originally in Brakel (2013), Diachronic Conjunctive Token Physicalism, in its brain-based view of mental content, could go some distance in explaining what the working-through mechanism might actually entail and therefore how working-through might actually work.

To review what I've said above (Section Two, *infra* pp.5–6), on the DiCoToP account, every singular mental event (including a specific mental content, e.g., a particular belief) exists synchronically (at a time) as a brain event consisting of an assembly of neurons, along with whatever neurochemical processes facilitate their connection. Further, each singular instance of the same event/same mental content (each token) is constituted by a slightly different population of neurons, a slightly different neuronal assembly,¹² insofar as it occurs at a different time and likely a different place. The sum of all of these instances of this content, i.e., all of their slightly variable neuronal networks (the conjunction), over time (diachronic), comprises the mental content.

¹²Sometimes the neuronal collections can show greater ensemble variation; indeed, even diverse neural circuits can at times be entrained.

Now, let's take a neurotic belief with a clear and simple mental content “All spiders are dangerous and should be avoided.” How would the working-through process actually occur according to the DiCoToP model of mental events? It seems clear that dealing with the phobic pathology that this specific mental content represents would require much re-aligning over a great number of neuronal assemblies, networks, and circuits, all summed over time. In other words, considerable experiential re-workings of the particular mental content would be needed—this, in most of the myriad contexts in which the neurotic belief appeared.

Working-through more complex neurotic symptoms, composed as they are of diverse and variegated neurotic beliefs, would clearly entail more realignments, more re-workings, more time, much of it taking place after gaining new knowledge from even the most successful interpretations. Diachronic Conjunctive Token Physicalism is then a mind/body view that can help explain not only working-through and the post-interpretative mechanism of therapeutic action, but also why analyses take so long, and why post analytic self-analysis is rarely optional.

Part Two: Can Neuroscientists Provide Evidence for (or Against) DiCoToP?

There is actual neuroscientific evidence for the sort of neuronal assemblies underlying mental processes that the Diachronic Conjunctive Token Physicalism (DiCoToP) model proposes (See Carillo-Reid and Yuste, 2020 for an overview). The mental processes include learning following sensory stimulus (Litwin-Kumar and Doiron, 2014); visual perception (Miller et al., 2017); tactile discrimination (Deolindo et al., 2017); appetitive learning (Brebner et al., 2020); formation of memories, concepts, words (Papadimitriou et al., 2020); pattern completion (Carillo-Reid et al., 2021); and learned reaction-time tasks (Narayanan et al., 2005). These studies all have shown (a) central stability in neuronal ensembles; and (b) ensembles that nonetheless include various levels of plasticity. These studies document that shifting neuronal membership occurs often, further demonstrating the type of redundancy seen in many biological systems (See Edelman and Gally, 2001, p.13766; Narayanan et al., 2005; Hiratani and Fukai, 2018). Moreover, these research findings have included degenerate neuronal arrays at the level of disparate circuits, reflecting an even greater degree of plasticity.

Note that in the experiments cited, the methods employed to study the neuronal assemblies—including their original formation, their variability, and resultant stability amidst the changes—have involved sophisticated two-photon calcium imaging/optogenetics and complex statistical modeling. With these tools, researchers documented important aspects of neuronal assemblies, including variability sometimes seen even in their synaptic connections,¹³ as well the more familiar specific neuronal membership shifts.

However, all of the optogenetic imaging and their encouraging results, were of necessity gathered on non-human animal subjects. What are the implications for human consciousness? The statistical modeling studies, on the other hand, were directly

¹³For synaptic redundancy, see Hiratani and Fukai (2018).

applied to human mental goings-on, and they are certainly informative. Returning to the animal studies, the subjects were clearly conscious in their responses to perceptual stimuli and various learning tasks. This means that the fine-grained (neuronal assembly, neuron, and even synapse-level) imaging results do reflect mental activity, which should be generalizable to human animals. And yet. . .

Here is one possibility offered by this author (a not-even-neophyte neuroscientist; but a psychoanalyst/philosopher) in order to test DiCoToP more specifically. Could the research methods described above (optogenetic imaging) be employed to study particular mental contents more directly in appropriate non-human animals? For example, two simple belief-like attitudes for a mouse might be selected: (1) “that is frightening;” and (2) “that is appealing.” Then, many versions of each of those content-rich beliefs would be delivered to each mouse-subject (e.g., 10 different mouse-subjects) in many iterations (20) during a single session, and then in succeeding sessions over some duration of time (e.g., 2 weeks)? Would imaging results compared within subjects and between subjects—both with regard to trials at-a-time, and those over-time—prove revealing?

And now, for the real point of this section, really of this whole article: I wrote it in the hopes of providing a challenge and registering a plea to neuroscience colleagues to devise human experiments offering empirical brain-based (perhaps neuronal assembly and neuron level) evidence for or against DiCoToP (as well as other competing mind/body views).

DISCUSSION: SUMMARY AND CONCLUSION

Any investigation of the perpetually vexing mind/body problem should not exclude a philosophical exploration, especially in light of the claim that neuroscientists and psychoanalysts hold clashing positions inhibiting their interdisciplinary collaboration. Thus, this article adds the philosophy of mind to the endeavor set forth as the “Frontiers Research Topic: Psychodynamic Neuroscience.” This piece begins with a series of questions: Why should one be a physicalist, when dualism can seem so right? Why should one be a reductive physicalist, when non-reductive physicalism has

the appeal of dualism, without some of its major problems? And finally, why token physicalism, when type physicalism has been much more popular?

Asking and answering these questions provides a review of some of the most important aspects of the mind/body dilemma. Among the issues taken up was the matter of independent causation. If the mind and the mental are properly reduced to the physical, how can mental goings-on have real autonomous, downward causal power? Are they not just epiphenomena? Emergence as a possible solution was discussed, along with the problems it does not resolve. Next, multiple realization, a concept that is truly pivotal for neuroscience and mind/body physicalists, was explored, and found to be particularly relevant as it has become a central feature in the mind/body view I developed in 2013, and outlined in the current article. Termed Diachronic Conjunctive Token Physicalism (DiCoToP), this account can resolve many of the outstanding mind/body problems. True, the DiCoToP account also brought to the fore a few new vexing philosophical puzzles. But after wrestling with these, the article took a more practical turn.

Several empirical studies, of two sorts—statistic modeling of neurological phenomena at the level of the neuron and neuronal assembly; and imaging studies on non-human animal subjects on fine-grained brain processes—were seen to provide evidence for essential aspects advanced by the DiCoToP account. This article ends with a challenge that is at the same time an appeal: Can neuroscientists devise empirical experiments with human subjects in order to gain evidence for (or against) the Diachronic Conjunctive Token Physicalist mind/body account?

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Bechtel, W., and Mundale, J. (1999). Multiple realizability revisited: linking cognitive and neural states. *Philos. Sci.* 66, 175–207. doi: 10.1086/392683
- Brakel, L. A. W. (2010). *Unconscious Knowing and Other Essays in Psycho-Philosophical Analysis*. Oxford: Oxford University Press.
- Brakel, L. A. W. (2013). *The Ontology of Psychology: Questioning Foundations in the Philosophy of Mind*. Oxford: Oxford University Press.
- Brebnier, S., Ziminski, Z., Margetts-Smith, G., Sieburg, M., Reeve, H., Nowotny, T., et al. (2020). The emergence of a stable neuronal ensemble from a wider pool of activated neurons in the dorsal medial prefrontal cortex during appetitive learning in mice. *J. Neurosci.* 40, 395–410. doi: 10.1523/JNEUROSCI.1496-19.2019
- Carillo-Reid, L., and Yuste, R. (2020). *What Is a Neuronal Ensemble? Subject: Computational Neuroscience*. Oxford: Oxford University Press, doi: 10.1093/acrefore/9780190264086.013.298
- Carillo-Reid, L., Shutling, H., O’Neil, D., Taralova, E., Jebara, T., and Yuste, R. (2021) Identification of pattern completion neurons in neuronal ensembles using probabilistic graphic models. *J. Neurosci.* 41, 8577–8588. doi: 10.1523/JNEUROSCI.0051-21.2021
- Chalmers, D. (2010). *The Character of Consciousness*. Oxford: Oxford University Press.
- Deolindo, C., Kunicki, A., da Silva, M., Lima Brasil, F., and Moiolli, R. (2017). Neuronal assemblies evidence distributed interactions within a tactile discrimination task in rats. *Front. Neural Circuits* 11:114. doi: 10.3389/fncir.2017.00114
- Edelman, G., and Gally, J. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13763–13768.
- Figdor, C. (2010). Neuroscience and the multiple realization of cognitive functions. *Philos. Sci.* 77, 419–456. doi: 10.1086/652964
- Friston, K. (1997). Imaging cognitive anatomy. *Trends Cogn. Sci.* 1, 21–27.

- Friston, K., and Price, C. (2003). Degeneracy and redundancy in cognitive anatomy. *Trends Cog. Sci.* 7, 151–152. doi: 10.1016/s1364-6613(03)00054-8
- Hill, C. (1991). *Sensations: A Defense of Materialism*. Cambridge: Cambridge University Press.
- Hill, C. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philos. Stud.* 87, 61–85.
- Hiratani, N., and Fukai, T. (2018). Redundancy in synaptic connections enables neurons to learn optimally. *PNAS* 115, E6871–E6879. doi: 10.1073/pnas.1803274115
- Jackson, F. (1982). Epiphenomenal qualia. *Philos. Q.* 3, 127–136. doi: 10.1098/rstb.1998.0346
- Jackson, F. (2003). “Mind and illusion,” in *Minds and Persons*, ed. A. O’Hear (New York, NY: Cambridge University Press), 25–272.
- Kim, J. (1993). *Supervenience and Mind*. Cambridge: Cambridge University Press.
- Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
- Leuenberger, S. (2008). *Ceteris abstentibus* physicalism. *Oxford Stud. Metaphys.* 4, 145–170.
- Litwin-Kumar, A., and Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nat. Commun.* 5:5319. doi: 10.1038/ncomms6319
- Miller, J. K., Ayzenshtat, I., Carrillo-Reid, L., and Yuste, R. (2017). Visual stimuli recruit intrinsically generated cortical ensembles. *PNAS* 111, E4053–E4061.
- Nagel, T. (1998). Conceiving the impossible and the mind-body problem. *Philosophy* 73, 337–352. doi: 10.1017/s0031819198000035
- Narayanan, N., Kimchi, E., and Laubach, M. (2005). Redundancy and synergy of neuronal ensembles in motor cortex. *J. Neurosci.* 25, 4207–4216. doi: 10.1523/jneurosci.4697-04.2005
- Noppeney, U., Friston, K., and Price, C. (2004). Degenerate neuronal systems sustaining cognitive functions. *J. Anat.* 205, 433–442. doi: 10.1111/j.0021-8782.2004.00343.x
- Papadimitriou, C., Vempalab, S., Mitropolskya, D., Collinsa, M., and Maass, W. (2020). Brain computation by assemblies of neurons. *PNAS* 117, 14464–14472.
- Price, C., and Friston, K. (1997). Cognitive conjunction; a new approach to brain activation experiments. *NeuroImage* 5, 261–270.
- Price, C., and Friston, K. (2002). Degeneracy and cognitive anatomy. *Trends Cogn. Sci.* 6, 416–421. doi: 10.1016/s1364-6613(02)01976-9
- Putnam, H. (ed.). (1967/1975). “The nature of mental states,” in *Mind, Language, and Reality: Philosophical Papers*, Vol. 2, (Cambridge: Cambridge University Press), 429–440. Chapter 21. doi: 10.1017/cbo9780511625251.023
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Smart, J. J. (1959). Sensations and brain processes. *Philos. Rev.* 68, 141–156.
- Strawson, G. (1994). *Mental Reality*. Cambridge, MA: MIT Press.
- Strawson, G. (2006). *Real Materialism and Other Essays*. Oxford: Clarendon Press.
- Strawson, G. (2009). *Selves*. Oxford: Oxford University Press.
- Yablo, S. (1992). Mental causation. *Philos. Rev.* 101, 245–280.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Brakel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Affective Neuroscience of Sexuality: Development of a LUST Scale

Jürgen Fuchshuber^{1,2,3,4*}, Emanuel Jauk^{4,5,6}, Michaela Hiebler-Ragger^{1,4} and Human Friedrich Unterrainer^{1,3,4,7*}

¹ Center for Integrative Addiction Research, Grüner Kreis Society, Vienna, Austria, ² Department of Philosophy, University of Vienna, Vienna, Austria, ³ University Clinic for Psychiatry and Psychotherapeutic Medicine, Medical University of Graz, Graz, Austria, ⁴ Department of Psychology, University of Graz, Graz, Austria, ⁵ Clinical Psychology and Behavioral Neuroscience, Technische Universität Dresden, Dresden, Germany, ⁶ University Department of Medical Psychology and Psychotherapy, Medical University Graz, Graz, Austria, ⁷ Department of Religious Studies, University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:

Christoph Mathys,
Aarhus University, Denmark

Reviewed by:

Andrea Clarici,
University of Trieste, Italy
Geir Pedersen,
Oslo University Hospital, Norway

*Correspondence:

Jürgen Fuchshuber
a00913689@unet.univie.ac.at
Human Friedrich Unterrainer
human.unterrainer@univie.ac.at

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 12 January 2022

Accepted: 07 February 2022

Published: 28 February 2022

Citation:

Fuchshuber J, Jauk E,
Hiebler-Ragger M and Unterrainer HF
(2022) The Affective Neuroscience
of Sexuality: Development of a LUST
Scale.
Front. Hum. Neurosci. 16:853706.
doi: 10.3389/fnhum.2022.853706

Background: In recent years, there have been many studies using the Affective Neuroscience Personality Scales (ANPS) to investigate individual differences in primary emotion traits. However, in contrast to other primary emotion traits proposed by Jaak Panksepp and colleagues, there is a considerable lack of research on the LUST (L) dimension – defined as an individual's capacity to attain sexual desire and satisfaction – a circumstance mainly caused by its exclusion from the ANPS. Therefore, this study aims to take a first step toward the development of a standardized self-rate measurement for the L-disposition. For this purpose, two versions of the L-scales (L-12 and L-5) were developed and evaluated regarding reliability and aspects of validity.

Materials and Methods: After a pilot study ($N = 204$; female: 81%) with an initial 20-item pool item reductions were conducted. This led to the construction of a 12-item (L-12) version and a 5-item version (L-5), which were assessed in a second sample consisting of 371 German-speaking healthy adults (58.50% female) aged 18–69 years ($M = 28$; $SD = 9.75$). Aspects of external validity were assessed by investigation of correlations with the ANPS, psychiatric symptoms (Brief Symptom Inventory-18), attachment security (Adult Attachment Scales) and personality functioning (Operationalized Psychodynamic Diagnostics Structure Questionnaire). To evaluate structural validity, both L-scales were investigated via confirmatory factor analysis (CFA).

Results: Cronbach's α indicated excellent internal consistency regarding L-12 ($\alpha = 0.90$), while L-5 showed acceptable reliability ($\alpha = 0.82$). CFA of a bifactor model of the L-12 indicated excellent model fit. Moreover, an excellent model fit was observed regarding a single factor model of L-5. For both scales small to moderate positive correlations were observed with SEEKING, PLAY, and secure attachment, while they exhibited small to moderate negative correlations with SADNESS, insecure attachment, lower personality functioning, and increased psychiatric symptom load.

Conclusion: Both newly developed scales exhibit satisfying psychometric properties, indicating high reliability, good structural validity and plausible correlations with external criteria. Hence, this study poses an important step toward the operationalization of the LUST concept. However, more research is needed in particular with respect to the scale's external validity and its applicability in clinical populations.

Keywords: LUST, questionnaire development, factor analysis, primary emotions, Affective Neuroscience

INTRODUCTION

Affects are considered as the paramount motivational foundation of human behavior, contributing a significant proportion to the formation of the individuals' temperament and personality (Kernberg, 2015). In recent decades, several attempts were made to differentiate and categorize basic affective systems (e.g., Izard, 1992; Panksepp, 1998; Ekman, 1999; Krause, 2012). In contrast to more cognitively oriented emotion/affect models, Panksepp (1998) emphasized the role of subcortical brain areas in the emergence of primary emotions [in Affective Neuroscience (AN) theory, *affect* is usually used synonymous with the term *emotion*]. Based on ethological research, he distinguished seven primary emotion networks, neuroanatomically located between the Periaqueductal Gray (PAG) and the limbic forebrain (Panksepp, 1998; Panksepp and Biven, 2012). These cross-species primary-process systems are assumed to act as prototype emotional states, which can be evoked via artificially induced stimulation (Panksepp and Watt, 2011). The Affective Neuroscience Personality Scales (ANPS; Davis et al., 2003) aim to measure six dimensions of these primary emotion systems including SEEKING, ANGER, FEAR, PLAY, SADNESS, CARE, and assess an additional spirituality scale.

While also conceptualized as a primary emotion, LUST was excluded from operationalization in the ANPS as the authors assumed that it “seems less relevant to current conceptualizations of human personality and we also suspected that it may potentially be an affective factor that people would not wish to be frank about, and thus may contaminate frankness on the other scales” (Davis and Panksepp, 2011, p. 1949). Nevertheless, LUST – defined as the “systems that contribute distinctly to female and male sexuality and associated erotic feelings” (Panksepp, 2004, p. 52) – remains a pivotal part of AN and related neuropsychanalytic theory, involved in important aspects of etiological models concerning psychiatric disorders (Panksepp, 2004; Boeker et al., 2018). Accordingly, it appears sensible to construct a standardized measurement of this dimension in order to facilitate psychometrically oriented neuropsychanalytic research. This demand is further highlighted by recent progress in this field achieved via the usage of the ANPS concerning investigations of primary emotions and its relationship to personality and phenomena of clinical psychology (Marengo et al., 2021; Montag et al., 2021).

Yet, to the best of our knowledge, no publicly available LUST scale exists until this day. One of the few trackable references to the LUST concept in relation to the ANPS can be found in a conference paper by Lotter and Dauphin (2014), which investigated the relationship between the ANPS and

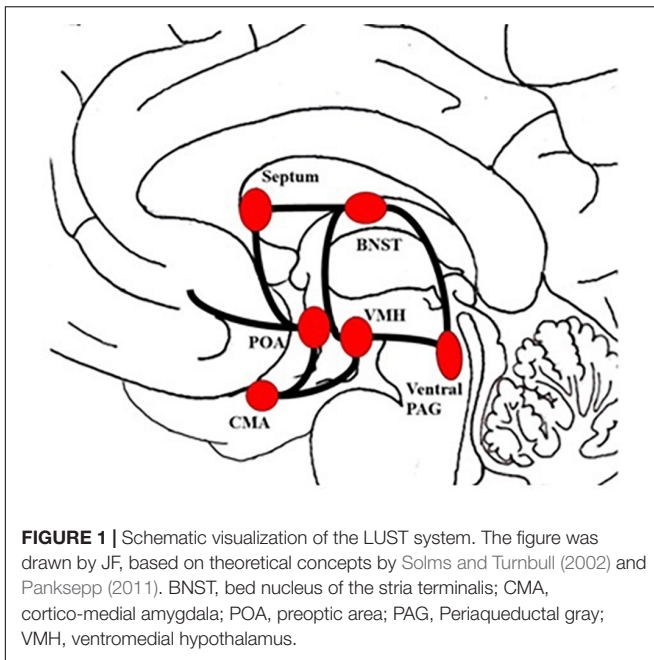
the Sexual Desire Inventory-2 (SDI-2; Spector et al., 1996), indicating a positive correlation between dyadic sexual desire and SEEK and PLAY as well as solitary and total sexual desire with SEEK and ANGER.

van der Westhuizen and Solms (2015) describe the development of a LUST measurement in context with the investigation of social dominance. According to the authors, the scale exhibited good internal consistency and was conceptualized via the assessment of “sexual desire, physiological arousal and sexual mentation.” However, this LUST scale was not made publicly available. Nevertheless, the authors reported moderate positive correlations between LUST and SEEKING and PLAY, small positive correlation with CARE and DOMINANCE as well as a small negative correlation with FEAR.

Another attempt to measure LUST was made by Dolce et al. (2020), who investigated a primary emotion driven personality concept by the means of neuronal networks training (Dolce et al., 2020). While the conceptual strategy of their newly developed self-rate measurement differs rather drastically from the original ANPS, the questionnaire primarily aims to achieve a high predictive validity with respect to psychiatric disorders. Therefore, in its current form, LUST is inversely assessed by three items which focus on inappetence, sadness and loneliness in social situations.

Based on its original definition, the LUST system is linked to feelings of pleasure, sexual urges and gratification, and is reciprocally connected to the SEEKING network (Panksepp, 1998; Solms and Turnbull, 2002; Panksepp and Biven, 2012). Its activation diminishes SEEKING driven appetite behavior and triggers feelings of satisfaction, serving as reward mediating substrate necessary for learning (Solms and Turnbull, 2002). This dichotomy between SEEKING and LUST largely corresponds to Robinson and Berridge's (2000) differentiation between “wanting” and “liking.” In correspondence to this, the process of learning is understood as a coupling between psychomotor SEEKING impulses (“wanting”) and the experience of pleasure mediated by the LUST system (“liking”).

Inferred from animal model, the LUST network is assumed to consist of a complex group of structures, descending from the hypothalamus to the posterior parts of the midbrain (see **Figure 1**; Panksepp, 1998; Solms and Turnbull, 2002). Most authors agree that the LUST system is composed of the Bed nucleus of the stria terminalis, the central tegmental field, the preoptic area and the ventromedial hypothalamus, the Nucleus accumbens shell, septum area and the ventral Periaqueductal gray (Solms and Turnbull, 2002; Holstege and Huynh, 2011; Panksepp and Biven, 2012; Berridge and Kringelbach, 2015; Hashikawa et al., 2016). Neurochemically,



the LUST system is largely controlled by endorphins acting on mu, delta- and kappa-opioid receptors in the Nucleus accumbens shell, and hormones like vasopressin, testosterone and oxytocin (Solms and Turnbull, 2011; Panksepp and Biven, 2012; Berridge and Kringelbach, 2015).

Moreover, recent studies in humans investigating the neural effects of orgasms with PET scans (Georgiadis et al., 2006; Holstege and Huynh, 2011) observed increased blood flow within the upper brainstem and cerebellum. Men showed increased activation in the insula, while women exhibited increased activation in the somatomotor and somatosensory cortex and a deactivation in the left temporal lobe and ventral prefrontal cortex.

Regarding the clinical significance of the LUST system, Panksepp (2004) exclusively proposed associations with disorders directly related to sexuality like sexual dysfunctions, sexual addictions, fetishes, and excessive jealousy. However, with regard to empirical research it seems plausible that dysfunctions of the LUST-system might be involved in a broad range of psychopathologies like affective disorders (Rokach, 2019), addictions (Kumsar et al., 2016), personality disorders (Collazzoni et al., 2017) and schizophrenia (Kelly and Conley, 2004). Yet, while there is growing evidence regarding associations between most primary emotions dispositions and psychopathology (Montag et al., 2021), due to the current lack of appropriate research instruments very little is known about the role of LUST.

As for associations with other personality traits, the current state of psychodynamic literature suggests associations between increased attachment security and personality functioning with a predominance of positive affective traits (Arbeitskreis OPD-2, 2008; Krause, 2012), an assumption supported by recent empirical evidence (Fuchshuber et al., 2019). Thus, it seems

plausible to assume that a measure of LUST is positively correlated with increased personality functioning and attachment security. However, with research investigating the role of attachment security in connection to the related concept of erotophilia [defined as positive respond and attitude bias toward sexual cues (Fisher et al., 1988)], indicating no significant correlations, this relationship seems less clear (Bogaert and Sadava, 2002). Despite conceptual overlaps between both erotophilia and a disposition toward LUST, the transferability of this result remains uncertain, as erotophilia might be more oriented toward conscious attitudes regarding sexuality than the personality trait LUST, defined by the individual's ability to attain sexual desire and pleasure.

Study Aims

This study aims to develop and psychometrically evaluate a preliminary measurement tool, which enables the assessment of the LUST concept decisively developed and described by Panksepp (1998). Hence, the scale might be seen as an exploratory addition to the already existing ANPS which – at its present form – do not operationalize LUST. In correspondence to this, we examine the reliability as well as aspects of the validity of this newly developed self-rate measurement.

Thereby, we expected the L-scales to correlate positively with adaptive attachment dimensions (trust and closeness) and negatively with attachment anxiety, as well as positively with personality functioning, and negatively with psychiatric symptom load.

MATERIALS AND METHODS

Item Generation

After an extensive literature search on available discourse on LUST as well as existing measurements of sexuality, the authors (JF, EJ, and MH-R) developed an item pool of 20 questions, which aimed to operationalize the concept as outlined by Panksepp (1998), Solms and Turnbull (2002), and Panksepp and Biven (2012). Hence, the scale was conceptualized as a single factor measurement of the individual's capacity to attain sexual desire, enjoyment and pleasure. Emphasis was placed on comprehensibility of the items, subsequently double negations, foreign words or technical terms were avoided. Furthermore, items were formulated as short and unambiguous as possible (Bühner, 2011).

Half of the items were constructed as negatively scored items and formatted in line with the BANPS (Barrett et al., 2013). Therefore, a five-point Likert scale was used to assess the individual items, as a response format of five to seven levels is suggested to maximize the reliability of assessment tools (Bühner, 2011). “1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree” was chosen as the response format.

Item Reduction

The initial item pool was tested in a pilot study ($N = 204$; female: 81%). This initial version of the LUST-scale was estimated to have

a Cronbach's α of 0.70. To achieve a reliable, unidimensional and valid scale this version was trimmed by the examination of (1) item-total correlations, aiming to achieve a Cronbach's $\alpha > 0.90$ for the long and a Cronbach's $\alpha > 0.80$ for the short version, (2) Item difficulty, thereby it was aimed to achieve a mixed set of item difficulties with a predominance of moderately difficult items, (3) Exploratory factor analysis to test for one-dimensionality and (4) considerations regarding construct validity. The trimming steps were carried out iteratively and led to the construction of a 12-item version (L-12). Moreover, by further reducing the number of items a five-item version (L-5) was developed.

Sample and Procedure

The participants ($N = 371$; 58.50% female) were recruited through advertising on social networks and public announcements at the Karl-Franzens University of Graz and Medical University of Graz. Informed consent was acquired before each participant filled in the test form that included demographic questions (e.g., age, sex, education status, and occupation status), the newly developed LUST scale as well as standardized questionnaires described below. No recompense was provided. The data were acquired via the online-survey platform LimeSurvey. Participants were included if they were aged over 18, stated no history of psychiatric disorders and completed all questionnaires.

The study was carried out in accordance with the Declaration of Helsinki. Ethical approval was granted by the Ethics Committee of the Medical University of Graz, Austria.

Psychometric Assessment

The German version of the *Affective Neuroscience Personality Scales* [ANPS; Davis et al., 2003; German version by Reuter and Henning (2015)] is a self-report measure which assesses behavioral traits associated with Panksepp's (1998) concept of primary emotion circuits. It consists of 110 items rated on a 4-point Likert scale ranging from 1 ("strongly disagree") to 4 ("strongly agree"). The scales of the ANPS include "SEEKING," "SADNESS," "FEAR," "ANGER," "CARE," and "PLAY" as well as an additional scale for spirituality.

The *Adult Attachment Scale* (AAS; Collins and Read, 1990; German version: Schmidt et al., 2004), which assesses the subject's anxiety about being rejected or unloved ("Anxiety"); comfort with closeness ("Close") and comfort with depending on others ("Depend"). The German version consists of 15 items (five items per scale), which are rated on a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").

The *Operationalized Psychodynamic Diagnostics Structure Questionnaire* (OPD-SQS; Ehrenthal et al., 2015) is a self-report measurement, which assesses deficits in personality functioning as proposed in the Operationalized Psychodynamic Diagnostic (Arbeitskreis OPD-2, 2008). This measure is comprised of 91 items which are rated on a 5-point Likert scale ranging from 0 ("strongly disagree") to 4 ("strongly agree"). The total score indicates deficits in overall personality functioning, with higher scores indicating more severe impairments.

The *Brief Symptom Inventory* (BSI-18; Derogatis, 2001; German version: Spitzer et al., 2011) is comprised of 18 items

which measure the amount of symptom burden in the last 7 days with regards to depression, anxiety and somatization. Items are rated on a 5-point Likert scale ranging from 0 "absolutely not" to 4 "very strong." The total score "Global Severity Index" can be generated by adding the scores of every item.

Statistical Analysis and Analysis Strategy

The confirmatory factor analysis (CFA) was conducted with AMOS 26. SPSS 27.0 was used for data management, descriptive statistics and bivariate correlations. Goodness-of-fit was assessed with maximum likelihood (ML) estimation in AMOS. In accordance with Kline (2015), the following global fit-indices were considered as markers for an acceptable model fit: (a) The Comparative Fit Index (CFI) > 0.90 ; (b) Tucker-Lewis Index (TLI) > 0.90 ; (c) the Normed Fit Index (NFI) > 0.90 ; (d) the Relative Fit Index (RFI) > 0.90 ; (e) the square root error of approximation (RMSEA) < 0.08 and the upper bound of its 90% confidence interval < 1 ; and (f) $\chi^2/df < 3$.

For the comparison of competing models, the Akaike Information Criterion (AIC) was used, which rewards models that achieve a high goodness-of-fit and penalizes them if they become overly complex (Kline, 2015). In this context, the model with the smallest AIC value was preferred, with a $\Delta AIC > 2$ indicating significant differences (Cheung and Rensvold, 2002; Jovanović, 2015).

To establish model identification one factor loading was fixed to 1 for each factor in every specified model (Byrne, 2004).

RESULTS

Sample Characteristics and Descriptive Statistics

The investigated sample consisted of 371 German-speaking adults (58.50% female) which were aged 18–69 years ($M = 28$; $SD = 9.75$). Most participants (39%) had a general qualification for university entrance as highest finished education, were in education (53%) and Austrian (80%). **Table 1** displays the sample characteristics in detail. Item characteristics of L-12 and L-5 are shown in **Table 2**. Item-total correlation ranged from $r_{iT} = 0.54$ – 0.68 for both scale versions. Item difficulty ranged from 45.22 to 63.07 in L-12 and from 45.22 to 57.01 in L-5.

With regard to sex differences, results indicated higher L-12 scores in males [$F_{(1,369)} = 6.49$; $p < 0.05$; $M_{male} = 3.25$; $SD_{male} = 0.56$; $M_{female} = 3.09$; $SD_{female} = 0.60$] with a small effect size of $\eta^2 = 0.02$, and no differences in L-5 [$F_{(1,369)} = 0.71$; $p > 0.05$].

Normal Distribution

Inspection of skewness and kurtosis indicated a normal distribution of both L-12 (skewness = 0.82; kurtosis = 0.74) and L-5 (skewness = -0.67 ; kurtosis = 0.22).

External Validity

L-12 and L-5 exhibited a large correlation ($r = 0.92$; $p < 0.001$; see Cohen, 1992). Consequently, both scales showed a similar

TABLE 1 | Sample characteristics ($N = 371$).

Sample	<i>N</i> or <i>M</i>	% or <i>SD</i>
Gender	<i>N</i> = 217 Female/ <i>N</i> = 154 Male	58.50% 41.50%
Age	<i>M</i> = 28	<i>SD</i> = 9.75 years
Highest finished education	<i>N</i> = 30 High School/ <i>N</i> = 146 General qualification for university entrance/ <i>N</i> = 72 Bachelor University degree/ <i>N</i> = 68 Master University degree/ <i>N</i> = 44 Apprenticeship/ <i>N</i> = 11 Ph.D./	8.10% 39.35% 19.41% 18.32% 11.86% 2.96%
Occupation	<i>N</i> = 150 In employment/ <i>N</i> = 198 In education/ <i>N</i> = 18 Unemployed/ <i>N</i> = 5 in Pension	40.43% 53.34% 4.85% 1.35%
Relationship Status	<i>N</i> = 45 Married/ <i>N</i> = 162 In Relationship/ <i>N</i> = 164 Single/	12.12% 43.67% 44.20%
Nationality	<i>N</i> = 291 Austrian/ <i>N</i> = 79 German/ <i>N</i> = 11 Other	80.05% 21.29% 2.96%

correlation pattern with external variables (see **Table 3**). While neither exhibited a significant relationship with ANGER and Spirituality (both $p > 0.05$), small to moderate positive correlations were observed with SEEKING, PLAY, Trust, and Closeness ($r = 0.28$ – 0.35 ; all $p < 0.001$). In contrast, L-12 and L-5 exhibited small to moderate negative correlations with SADNESS, and anxiety about being rejected or unloved,

personality functioning deficits, and psychiatric symptom burden ($r = -0.25$ to -0.39 ; all $p < 0.001$). CARE showed a small positive correlation with L-5 ($r = 0.14$; $p < 0.01$) but not L-12 ($p > 0.05$).

Reliability

Cronbach's α indicated excellent internal consistency regarding L-12 ($\alpha = 0.90$), while L-5 showed acceptable reliability ($\alpha = 0.82$).

Confirmatory Factor Analyses of the LUST-Scale

Confirmatory factor analysis fit indices for the 12-item and 5-item version are detailed in **Table 4**. An initial solution of the 12-item version without correlated error terms estimated a poor fitting model: $\chi^2/df = 8.02$; $RMSEA = 0.14$ (90% CI: 0.13, 0.15); $CFI = 0.81$; $NFI = 0.79$; $TLI = 0.77$; $RFI = 0.74$.

Based on the inspection of error term correlations, a bifactor model of the L-12 scale was investigated, which specified a general factor (L-12) and two residual factors modeled via the assignments of (1) positively poled items and (2) negatively poled items. Results indicated an excellent model fit: $\chi^2/df = 2.02$; $RMSEA = 0.05$ (90% CI: 0.04, 0.70); $CFI = 0.97$; $NFI = 0.96$; $TLI = 0.97$; $RFI = 0.93$.

L-5 showed an excellent fit with the following indices: $\chi^2/df = 0.56$; $RMSEA = 0.00$ (90% CI: 0.00, 0.05); $CFI = 1.00$; $NFI = 1.00$; $TLI = 1.01$; $RFI = 0.99$.

As detailed in **Figure 2**, the factor loadings of the initial L-12 model ranged from $\beta = 0.53$ to $\beta = 0.73$ (all $p < 0.001$). Regarding the bifactor model all associations with the total factor ranged from $\beta = 0.53$ to $\beta = 0.81$, while significant associations with the positive item residual factor were estimated between $\beta = 0.24$

TABLE 2 | Item-characteristics of L-12 and L-5 scales.

Scale and Items	$\bar{X}_i$	S_i	r_{IT}	P_i
L-12				
1. Mir fällt es leicht, mich erotischen Erfahrungen hinzugeben	2.81	0.89	0.63	45.22
2. Meine Sexualität auszuleben, fühlt sich irgendwie nicht richtig an*	3.27	0.87	0.57	56.67
3. Ich empfinde meine Sexualität allgemein als befriedigend	2.93	0.86	0.62	48.32
4. Sexualität ist für mich mit Ekel verknüpft*	3.52	0.77	0.65	63.07
5. Mir fällt es leicht, einen Orgasmus zu haben	2.95	0.89	0.53	48.72
6. Ich stehe Sexualität nicht besonders offen gegenüber*	3.19	0.88	0.63	54.78
7. Ich kann das Ausüben von sexuellen Handlungen (Geschlechtsverkehr, Masturbation, etc.) voll und ganz genießen	3.25	0.79	0.73	56.20
8. Mit meiner Sexualität habe ich häufig schlechte Erfahrungen gemacht*	3.28	0.84	0.58	57.01
9. Wenn ich Sex habe, komme ich meist zu einem Orgasmus	3.04	0.99	0.57	50.94
10. Den Anblick von Geschlechtsorganen finde ich abstoßend*	3.42	0.80	0.60	60.51
11. Wenn ich einen Orgasmus habe, nehme ich ihn meistens sehr intensiv wahr	2.99	0.84	0.54	49.87
12. Sexualität ist für mich sehr mit Scham besetzt*	3.22	0.89	0.68	55.59
L-5				
1. Mir fällt es leicht, mich erotischen Erfahrungen hinzugeben	2.81	0.89	0.66	45.22
2. Mit meiner Sexualität habe ich häufig schlechte Erfahrungen gemacht*	3.28	0.84	0.54	57.01
3. Ich empfinde meine Sexualität allgemein als befriedigend	2.93	0.86	0.64	48.32
4. Ich stehe Sexualität nicht besonders offen gegenüber*	3.19	0.88	0.54	54.78
5. Ich kann das Ausüben von sexuellen Handlungen (Geschlechtsverkehr, Masturbation, etc.) voll und ganz genießen	3.25	0.79	0.68	56.20

$N = 371$; $\bar{X}_i$, Mean of Item scores; S_i , Standard Deviation of Item scores; r_{IT} , Item Total Correlation; P_i , Item Difficulty. *Items are scored inversely.

TABLE 3 | Descriptive statistics, internal consistency and correlations between the L-scales and measures of personality structure as well as psychiatric symptom burden.

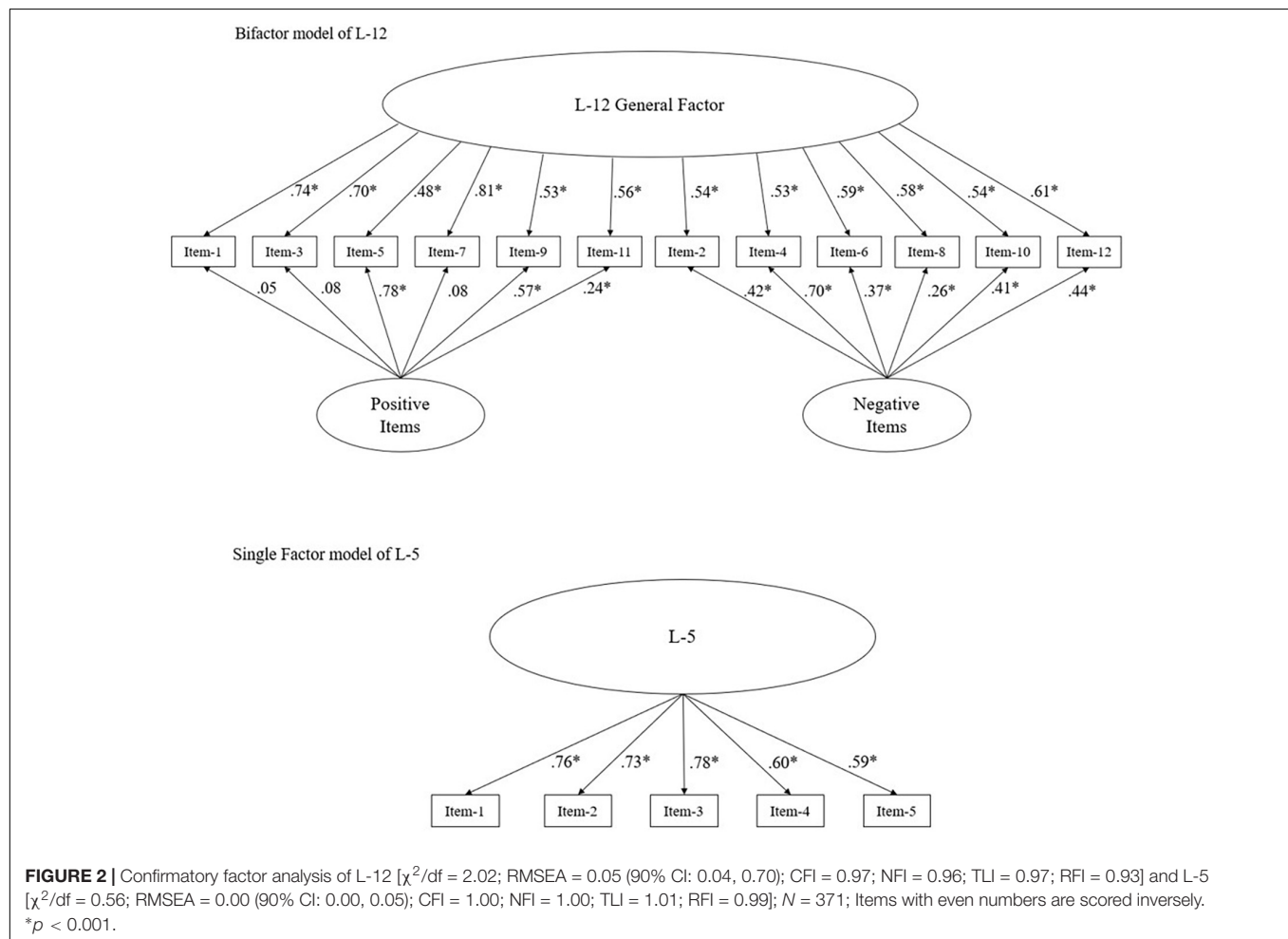
Measurement	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
LUST	1. L-12	–																
	2. L-5	0.93**	–															
ANPS	3. SEEKING	0.30**	0.32**	–														
	4. FEAR	–0.30**	–0.31**	–0.24**	–													
	5. CARE	0.13	0.14*	0.24**	0.14*	–												
	6. PLAY	0.32**	0.35**	0.50**	–0.30**	0.39**	–											
	7. ANGER	–0.04	–0.06	–0.02	0.33**	–0.03	–0.11	–										
	8. SADNESS	–0.25**	–0.27**	–0.25**	0.68	0.08	–0.29**	0.37**	–									
	9. Spirituality	0.05	0.05	0.13	–0.06	0.13	0.15*	–0.06	0.01	–								
BSI-18	10. Somatization	–0.21**	–0.21**	–0.10	0.31**	0.03	–0.17**	0.23**	–0.03	0.30**	–							
	11. Depression	–0.29**	–0.33**	–0.30**	0.50**	–0.07	–0.34**	0.21**	–0.03	0.59**	0.48**	–						
	12. Anxiety	–0.20**	–0.22**	–0.10	0.48**	0.03	–0.17**	0.27**	–0.01	0.45**	0.63**	0.63**	–					
	13. GSI	–0.28**	–0.31**	–0.21**	0.52**	–0.01	–0.28**	0.28**	–0.03	0.55**	0.79**	0.87**	0.88**	–				
OPD	14. PS	–0.37**	–0.39**	–0.28**	0.55**	–0.10	–0.37**	0.25**	0.09	0.54**	0.52**	0.63**	0.58**	0.69**	–			
AAS	15. Trust	0.28**	0.29**	0.25**	–0.31**	0.21**	0.45**	–0.29**	0.07	–0.40**	–0.36**	–0.54**	–0.35**	–0.51**	–0.54**	–		
	16. Closeness	0.37**	0.38**	0.28**	–0.26**	0.25**	0.50**	–0.14	0.11	–0.25**	–0.33**	–0.38**	–0.30**	–0.40**	–0.56**	0.52**	–	
	17. Anxiety	–0.27**	–0.28**	–0.19**	0.53**	0.09	–0.20**	0.22**	0.00	0.64**	0.34**	0.55**	0.52**	0.57**	0.66**	–0.44**	–0.27**	–
Descriptive	M	3.16	3.09	2.89	2.64	2.90	2.90	2.53	2.30	2.52	9.57	11.19	1.71	31.48	64.43	15.94	13.24	11.06
	SD	0.59	0.65	0.37	0.52	0.41	0.45	0.48	0.62	0.41	3.79	5.37	4.00	11.17	18.37	4.57	4.75	4.43
	α	0.90	0.82	0.74	0.87	0.75	0.82	0.84	0.87	0.74	78	0.88	0.80	0.91	0.96	0.82	0.86	0.77

N = 371; ANPS, Affective Neuroscience Personality Scales; GSI, BSI-18 Total Score; PS, OPD Personality Structure; M, mean; SD, standard deviation α , Cronbach's α . **p* < 0.01; ***p* < 0.001.

TABLE 4 | Results of the confirmatory factor analysis.

Model	χ^2 (df)	χ^2/df	RMSEA (90% CI)	CFI	NFI	TLI	RFI
12-Item version (without error term correlation)	433.13 (54)	8.02	0.138 (0.128–0.150)	0.810	0.790	0.768	0.744
12-Item version (Bifactor model)	84.22 (42)	2.01	0.053 (0.036–0.069)	0.979	0.959	0.967	0.936
5-Item version	2.79 (5)	0.56	0.000 (0.000–0.052)	1.000	0.995	1.007	0.991

$N = 371$.



and $\beta = 0.78$ (all $p < 0.001$) and between $\beta = 0.26$ to $\beta = 0.70$ for the negative item residual factor (all $p < 0.001$). Three items did not show significant association with either residual factor (all $p > 0.05$).

Results for the five-item version indicated factor loadings ranging from $\beta = 0.53$ to $\beta = 0.76$ (all $p < 0.001$).

DISCUSSION

This study aimed to be a building block toward the development of a self-rate measurement for the operationalization of LUST as proposed by Panksepp (1998). The psychometric properties presented in this paper suggest high reliabilities, normal distribution, good structural validity and a nomological network

conforming to our hypotheses of both the long- and short-version of the proposed L-scales.

With respect to reliability, L-12 and L-5 exhibited excellent and good internal consistency, respectively, with comparable or even higher Cronbach's α than the scales of the ANPS. For future studies further markers of reliability (e.g., split half or retest reliability) should be estimated. Additionally, inspection of skewness and kurtosis suggested a normal distribution of both scales with regard to a healthy population, again resonating with the characteristics of original ANPS (Davis et al., 2003; Davis and Panksepp, 2011).

Confirmatory factor analysis of L-12 revealed a bifactorial structure, with one general latent factor and two residual factors which seem to reflect specific variance regarding different answering styles with respect to items formulated positively

vs. items which assessed LUST inversely. Alternatively, this result might be interpreted as residual variance caused by items concerning orgasms and items operationalizing disgust and shame, respectively. However, this effect vanished in the short version. L-5 exhibited excellent model fit in a single factor model, making it especially useful for studies applying structural equation modeling.

As the original version of the ANPS is often considered too long, currently two brief versions of the questionnaire have been developed, namely the ANPS short (ANPS-S, Pingault et al., 2012), and the brief ANPS (BANPS, Barrett et al., 2013). According to Pedersen et al. (2014) short versions generally improve the psychometric properties of the ANPS, however, the authors abstained from constructing a corresponding LUST scales. In this context, the L-5 scale might prove itself as a practical supplement for both ANPS short versions. In correspondence to this, the present form of the L-Scales was constructed in accordance with 5-point response format of the BANPS and consequently differs from 4-point format of the original ANPS.

The construction of the L-scales was aimed to be closely aligned to the original definition of LUST as introduced by Panksepp (1998, 2012), hence the investigated items consist of statements associated with eroticism, sexual enjoyment and pleasure. Of note, for the short version of the questionnaire, all statements which mentioned orgasms had to be deleted in order to establish a robust single-factor structure. Considering significant sex differences in the frequency of orgasms during intercourse (Frederick et al., 2018), the exclusion of statements focused on orgasms might explain the disappearance of sex differences in the short version. Nevertheless, both L-12 and L-5 exhibit a strong intercorrelation and a very similar correlation pattern with external criteria.

In correspondence to this, the L-scales exhibited moderate positive correlations with dispositions toward SEEKING and PLAY, while there were only small to non-significant correlations with CARE. Furthermore, both scales showed moderate negative correlations with SADNESS and FEAR, whereas both were independent of ANGER. Accordingly, the L-scales appear to be more inversely related to negative primary emotions than the scale developed by van der Westhuizen and Solms (2015). Hence, the results resonate with current neuropsychodynamic literature, which assumes underlying drives (libidinal vs. destructive) acting as a latent grouping mechanism for affects (Krause, 2012; Kernberg, 2015; Boeker et al., 2018). However, further studies employing structural equation modeling which consider all seven primary emotions in a single model should be done to further investigate this hypothesis.

In accordance with the current psychodynamic understanding of the functional role of attachment and personality structure regarding affect regulation (Krause, 2012; Kernberg, 2015; Fonagy, 2018), the observed positive correlations between LUST and attachment security as well as more mature personality structure echo recent evidence indicating secure attachment and personality functioning to predict a pattern of increased positive and decreased negative primary emotions (Fuchshuber et al., 2019).

Along these lines, LUST appeared to be negatively related to increased psychiatric symptom burden. More detailed inspection suggests that this is predominantly driven by its negative association with depressive symptoms. Nevertheless, we also found significant (but small) negative correlations with symptoms of anxiety and somatization. These findings are in agreement with the current review of Rokach (2019), who detected that sexual dysfunctions are a wide spread concomitant of affective disorders. With regard to clinically relevant associations proposed by Panksepp (2004) future research should investigate the relationship between the L-scales and pathological jealousy and hypersexuality.

What is more, neither scale showed a significant correlation with spirituality as assessed with the ANPS. Relatively little is known about the relationship between LUST and spirituality, however, Brelsford et al. (2011) reported findings indicating a moderate negative correlation of sexuality and spirituality. Nevertheless, considering spiritual traditions which involve sexual practices (e.g., Tantra) further research may be needed to evaluate this relationship in more detail.

Limitations and Future Perspectives

Several limitations of this study must be noted which restrain the interpretation of the results and need to be addressed in future research.

A next step in the validation of the L-scales will be the evaluation of its external validity based on measures of sexuality like the Sexual Desire Inventory (SDI, Spector et al., 1996) or the Multidimensional Sexuality Questionnaire (Snell et al., 1993). While further convergent validity tests are needed, the L-scales presented here have high face validity (i.e., item content clearly assess the individual capacity to attain sexual pleasure), and results provide clear evidence for a common factor underlying all items, as well as a nomological network with respect to relevant aspects of personality which is in good accordance with our hypotheses.

Another critical issue of the current L-scales is that neither the long- nor the short-version take into account the frequency of sexual behavior as well as the aspect of sexual desire, as corresponding items were excluded based on initial item statistics and considerations regarding internal consistency. Nevertheless, sexual urges and the prevalence of sexual behavior are usually (but not always; see e.g., Solms and Turnbull, 2002) proposed as important markers for the expression of LUST (Panksepp and Biven, 2012). Therefore, a revised version of the L-scales should aim at reformulation and inclusion of associated items.

The present study is further limited due to the nature of its sample. The convenient sample was recruited online – an approach which is often controversially discussed in literature (see e.g., Wiersma, 2013) – and predominantly consisted of healthy students from Austrian universities. Thereby, the presence of psychiatric disorders was assessed via self-report and hence, there is a lack of systematic assessment by eligible diagnostic instruments (e.g., SKID, First et al., 2002). Along these lines, the L-scales needs to be evaluated in clinical populations in order to learn more about its etiological relevance.

Finally, our findings are restricted due to a lack of differentiation regarding the sexual orientation or gender identity of participants, as no plausible *a priori* reasons of significant differences were assumed.

Conclusion

This study aimed to complement the ANPS framework by a LUST scale – a concept which has been discussed as a vital element yet lacks a standardized assessment. The present data suggests that the L-scales scales are distinguished by high reliability, satisfying structural validity and plausible correlations with external criteria. Therefore, this study might serve as vital groundwork for a standardized operationalization of LUST. However, more research will be necessary to further evaluate this questionnaire, especially with considerations to external validity and its applicability in clinical populations.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Arbeitskreis OPD-2 (2008). *Operationalized Psychodynamic Diagnosis OPD-2: Manual of Diagnosis and Treatment Planning*. Göttingen: Hogrefe Publishing.
- Barrett, F. S., Robins, R. W., and Janata, P. (2013). A brief form of the affective neuroscience personality scales. *Psychol. Assess.* 25, 826–843. doi: 10.1037/a0032576
- Berridge, K. C., and Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron* 86, 646–664. doi: 10.1016/j.neuron.2015.02.018
- Boeker, H., Hartwich, P., and Northoff, G., eds (2018). *Neuropsychodynamic Psychiatry*. Luxemburg: Springer.
- Bogaert, A. F., and Sadava, S. (2002). Adult attachment and sexual behavior. *Pers. Relatsh.* 9, 191–204. doi: 10.1111/1475-6811.00012
- Brelsford, G. M., Luquis, R., and Murray-Swank, N. A. (2011). College students' permissive sexual attitudes: links to religiousness and spirituality. *Int. J. Psychol. Relig.* 21, 127–136. doi: 10.1080/10508619.2011.557005
- Bühner, M. (2011). *Einführung in Die Test-und Fragebogenkonstruktion*. Hallbergmoos: Pearson Deutschland GmbH.
- Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: a road less traveled. *Struct. Equ. Model.* 11, 272–300. doi: 10.1207/s15328007sem1102_8
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037//0033-2909.112.1.155
- Collazzoni, A., Ciocca, G., Limoncin, E., Marucci, C., Mollaioli, D., Di Sante, S., et al. (2017). Mating strategies and sexual functioning in personality disorders: a comprehensive review of literature. *Sex. Med. Rev.* 5, 414–428. doi: 10.1016/j.sxmr.2017.03.009
- Collins, N. L., and Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *J. Pers. Soc. Psychol.* 58, 644–663. doi: 10.1037//0022-3514.58.4.644
- Davis, K. L., and Panksepp, J. (2011). The brain's emotional foundations of human personality and the affective neuroscience personality scales. *Neurosci. Biobehav. Rev.* 35, 1946–1958. doi: 10.1016/j.neubiorev.2011.04.004
- Davis, K. L., Panksepp, J., and Normansell, L. (2003). The affective neuroscience personality scales: normative data and implications. normative data and

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical University of Graz. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JF conducted the all statistical analysis and wrote the first draft of the manuscript. EJ, MH-R, and HU read the manuscript and made some critical comments. JF, MH-R, EJ, and HU revised the whole manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.853706/full#supplementary-material>

- implications. *Neuropsychanalysis* 5, 57–69. doi: 10.1080/15294145.2003.10773410
- Derogatis, L. R. (2001). *Brief Symptom Inventory 18*. Baltimore: Johns Hopkins University.
- Dolce, P., Marocco, D., Maldonato, M. N., and Sperandio, R. (2020). Toward a machine learning predictive-oriented approach to complement explanatory modeling. an application for evaluating psychopathological traits based on affective neurosciences and phenomenology. *Front. Psychol.* 11:446. doi: 10.3389/fpsyg.2020.00446
- Ehrenthal, J. C., Dinger, U., Schauenburg, H., Horsch, L., Dahlbender, R. W., and Gierk, B. (2015). Entwicklung einer Zwölf-Item-Version des OPD-Strukturfragebogens (OPD-SFK)/Development of a 12-item version of the OPD-Structure Questionnaire (OPD-SQS). *Z. Psychosom. Med. Psychother.* 61, 262–274. doi: 10.13109/zptm.2015.61.3.262
- Ekman, P. (1999). “Basic emotions,” in *Handbook of Cognition and Emotion*, ed. T. Dalgleish and M. Power (Hoboken, NJ: John Wiley & Sons), 16.
- First, M. B., Spitzer, R. L., Gibbon, M., and Williams, J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York, NY: Biometrics Research.
- Fisher, W. A., White, L. A., Byrne, D., Kelley, K., and Gray, J. (1988). Erotophobia-erotophilia as a dimension of personality // Erotophobia-erotophilia and sexual behavior during pregnancy and postpartum. *J. Sex Res.* 25, 123–151. doi: 10.1080/00224498809551469
- Fonagy, P. (2018). *Affect Regulation, Mentalization and the Development of the Self*. Abingdon: Routledge.
- Frederick, D. A., John, H. K. S., Garcia, J. R., and Lloyd, E. A. (2018). Differences in orgasm frequency among gay, lesbian, bisexual, and heterosexual men and women in a US national sample // Differences in Orgasm Frequency Among Gay, Lesbian, Bisexual, and Heterosexual Men and Women in a U.S. National Sample. *Arch. Sex. Behav.* 47, 273–288. doi: 10.1007/s10508-017-0939-z
- Fuchshuber, J., Hiebler-Ragger, M., Kresse, A., Kapfhammer, H.-P., and Unterrainer, H. F. (2019). The influence of attachment styles and personality organization on emotional functioning after childhood trauma. *Front. Psychiatry* 10:643. doi: 10.3389/fpsyg.2019.00643
- Georgiadis, J. R., Kortekaas, R., Kuipers, R., Nieuwenburg, A., Pruim, J., Reinders, A. A. T. S., and Holstege, G. (2006). Regional cerebral blood flow changes associated with clitorally induced orgasm in healthy women. *Eur. J. Neurosci.* 24, 3305–3316. doi: 10.1111/j.1460-9568.2006.05206.x

- Hashikawa, K., Hashikawa, Y., Falkner, A., and Lin, D. (2016). The neural circuits of mating and fighting in male mice. *Curr. Opin. Neurobiol.* 38, 27–37. doi: 10.1016/j.conb.2016.01.006
- Holstege, G., and Huynh, H. K. (2011). Brain circuits for mating behavior in cats and brain activations and de-activations during sexual stimulation and ejaculation and orgasm in humans. *Horm. Behav.* 59, 702–707. doi: 10.1016/j.yhbeh.2011.02.008
- Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol. Rev.* 99, 561–565. doi: 10.1037/0033-295X.99.3.561
- Jovanović, V. (2015). Structural validity of the mental health continuum-short form: the bifactor model of emotional, social and psychological well-being. *Pers. Individ. Differ.* 75, 154–159. doi: 10.1016/j.paid.2014.11.026
- Kelly, D. L., and Conley, R. R. (2004). Sexuality and schizophrenia: a review. *Schizophr. Bull.* 30, 767–779. doi: 10.1093/oxfordjournals.schbul.a007130
- Kernberg, O. F. (2015). Neurobiological correlates of object relations theory. The relationship between neurobiological and psychodynamic development. *Int. Forum Psychoanal.* 24, 38–46. doi: 10.1080/0803706X.2014.912352
- Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Publications.
- Krause, R. (2012). *Allgemeine Psychodynamische Behandlungs- und Krankheitslehre: Grundlagen und Modelle*. Stuttgart: Kohlhammer Verlag.
- Kumsar, N. A., Kumsar, S., and Dilbaz, N. (2016). Sexual dysfunction in men diagnosed as substance use disorder. *Andrologia* 48, 1229–1235. doi: 10.1111/and.12566
- Lotter, H., and Dauphin, B. (2014). “What ever happend to LUST? Expanding the measures of primary process subcortical brain systems for the affective neuroscience personality scales,” in *Proceedings of the Society for Personality Assessment 2014*, Arlington, VA.
- Marengo, D., Davis, K. L., Gradwohl, G. Ö, and Montag, C. (2021). A meta-analysis on individual differences in primary emotional systems and Big Five personality traits. *Sci. Rep.* 11:7453. doi: 10.1038/s41598-021-84366-8
- Montag, C., Elhai, J. D., and Davis, K. L. (2021). A comprehensive review of studies using the affective neuroscience personality scales in the psychological and psychiatric sciences. *Neurosci. Biobehav. Rev.* 125, 160–167. doi: 10.1016/j.neubiorev.2021.02.019
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford university press.
- Panksepp, J. (2004). *Textbook of Biological Psychiatry*. Hoboken, NJ: John Wiley & Sons.
- Panksepp, J. (2011). The basic emotional circuits of mammalian brains: do animals have affective lives? *Neurosci. Biobehav. Rev.* 35, 1791–1804. doi: 10.1016/j.neubiorev.2011.08.003
- Panksepp, J. (2012). What is an emotional feeling? Lessons about affective origins from cross-species neuroscience. *Motiv. Emot.* 36, 4–15.
- Panksepp, J., and Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York, NY: WW Norton & Company.
- Panksepp, J., and Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emot. Rev.* 3, 387–396. doi: 10.1177/1754073911410741
- Pedersen, G., Selsbakk, J. M., Theresa, W., and Sigmund, K. (2014). Testing different versions of the affective neuroscience personality scales in a clinical sample. *PLoS One* 9:e109394. doi: 10.1371/journal.pone.0109394
- Pingault, J.-B., Falissard, B., Côté, S., and Berthoz, S. (2012). A new approach of personality and psychiatric disorders: a short version of the affective neuroscience personality scales. *PLoS One* 7:e41489. doi: 10.1371/journal.pone.0041489
- Reuter, M., and Henning, J. (2015). *Affective Neuroscience Personality Scales (ANPS). German Version*. Giessen: University of Giessen. Unpublished manuscript.
- Robinson, T. E., and Berridge, K. C. (2000). The psychology and neurobiology of addiction: an incentive-sensitization view. An incentive-sensitization view. *Addiction* 95 Suppl 2, S91–S117. doi: 10.1080/09652140050111681
- Rokach, A. (2019). The effect of psychological conditions on sexuality: a review. *Psychol. Psychother. Res. Study* 2, 259–266. doi: 10.31031/PPRS.2019.02.000534
- Schmidt, S., Strauss, B., Höger, D., and Brähler, E. (2004). The Adult Attachment Scale (AAS)-psychometric evaluation and normation of the German version. *Psychother. Psychosom. Med. Psychol.* 54, 375–382. doi: 10.1055/s-2003-815000
- Snell, W. E., Fisher, T. D., and Walters, A. S. (1993). The multidimensional sexuality questionnaire: an objective self-report measure of psychological tendencies associated with human sexuality. *Ann. Sex Res.* 6, 27–55. doi: 10.1007/BF00849744
- Solms, M., and Turnbull, O. (2002). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. London: Karnac Books.
- Solms, M., and Turnbull, O. H. (2011). What is neuropsychanalysis? *Neuropsychanalysis* 13, 133–145. doi: 10.1080/15294145.2011.10773670
- Spector, I. P., Carey, M. P., and Steinberg, L. (1996). The sexual desire inventory: development, factor structure, and evidence of reliability. *J. Sex Marital Ther.* 22, 175–190. doi: 10.1080/00926239608414655
- Spitzer, C., Hammer, S., Löwe, B., Grabe, H. J., Barnow, S., Rose, M., et al. (2011). Die Kurzform des Brief Symptom Inventory (BSI –18): erste Befunde zu den psychometrischen Kennwerten der deutschen Version. Erste Befunde zu den psychometrischen Kennwerten der deutschen Version [The short version of the Brief Symptom Inventory (BSI –18): preliminary psychometric properties of the German translation]. *Fortschr. Neurol. Psychiatr.* 79, 517–523. doi: 10.1055/s-0031-1281602
- van der Westhuizen, D., and Solms, M. (2015). Social dominance and the affective neuroscience personality scales. *Conscious Cogn.* 33, 90–111. doi: 10.1016/j.concog.2014.12.005
- Wiersma, W. (2013). The validity of surveys: online and offline. *Oxford Internet Inst.* 18, 321–340.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fuchshuber, Jauk, Hiebler-Ragger and Unterrainer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Memory for the Future: Psychodynamic Approach to Time and Self Through the Default Network

Filippo Cieri*

Department of Neurology, Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, United States

OPEN ACCESS

Edited by:

Lutz Jäncke,
University of Zurich, Switzerland

Reviewed by:

Alexander Fingelkurts,
BM-Science, Finland
Jeremy Holmes,
University of Exeter, United Kingdom

*Correspondence:

Filippo Cieri
filippocieri@gmail.com

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 27 February 2022

Accepted: 20 April 2022

Published: 16 June 2022

Citation:

Cieri F (2022) Memory
for the Future: Psychodynamic
Approach to Time and Self Through
the Default Network.
Front. Hum. Neurosci. 16:885315.
doi: 10.3389/fnhum.2022.885315

Time exists in us, and our self exists in time. Our self is affected and shaped by time to the point that a better understanding of the former can aid the understanding of the latter. Psychoanalysis works through self and time, where the self is composed of the biopsychosocial history (the past) of the individual and able to map a trajectory for the future. The psychoanalytic relationship starts from a “measurement”: an active process able to alter the system being measured—the self—continuously built over time. This manuscript, starts from the philosophical and scientific tradition of a proximity between time and self, suggesting a neural overlapping at the Default Network. A historical and scientific background will be introduced, proposing a multidisciplinary dimension that has characterized the birth of psychoanalysis (its past), influencing its present and future in the dialogue with physics and neuroscience. After a historical scientific introduction, a neural *entanglement* between past and future at the Default Network level will be proposed, tracing a link with the self at the level of this network. This hypothesis will be supported by studies in cognitive neurosciences and functional neuroimaging which have used the resting state functional Magnetic Resonance Imaging. The ontogenetic development of time perception will be discussed, consistent with self-development and the Default Network's function. The most common form of dementia, the Alzheimer's Disease, in which the perception of time is brutally impaired together with a loss of the self's functions will be proposed to support this idea. Finally, the potential theoretical and clinical significance for psychoanalysis and psychodynamic neurosciences, will be discussed.

Keywords: memory, time, default network, psychodynamic neuroscience, brain imaging, resting state—fMRI, psychoanalysis

“Who controls the past controls the future: who controls the present controls the past.”

(Nineteen Eighty-Four; Orwell, 1949).

INTRODUCTION

The proximity between time and self has deep philosophical roots. The experience of time is made by the self, where subjective time is associated to the conscious self as an enduring entity over time (Edmund Husserl) and subjective time being in essence embodied (Maurice Merleau-Ponty) (Zahavi, 2005; Wittmann, 2009, 2013). Recent neuroscientific approaches have underlined

the affective and interoceptive states of the body as necessary to create a sense of time (Craig, 2009, 2015; Droit-Volet and Gil, 2009; Wittmann et al., 2015).

The renewed discussion of the free energy principle, has given to the *mindbrain* system a greater probabilistic perspective, restarting a dialogue among physics, philosophy, psychology, neuroscience, and psychoanalysis that began with Freud. In this dialogue with neuroscience, neuropsychanalysis has been accused of using the chemistry of ink to appreciate Van Gogh's art (Blass and Carmeli, 2007; see Yovell et al., 2015 for a reply to this criticism). In other words, the neuroscientific approach in psychoanalysis has been accused of using the brain's language to translate the mind's dimension. Although neuroscience and psychoanalysis have different languages, often studying different aspects of the mindbrain, a dual-aspect monistic position on the philosophical mind-body problem allows for the investigation of the causal mechanism of consciousness not in the manifest brain but rather in its functional organization, which ultimately underpins both the physiological and the psychological manifestations of experience (Cieri and Esposito, 2019). A dual-aspect monism approach will be used to explore the relationship between self and time in this manuscript. Spinoza, Fechner, Schopenhauer, all shared a dual-aspect monistic view, and one of the most intriguing interdisciplinary contributions in this perspective comes from the cooperation between physics and psychoanalysis personified by the collaboration between Wolfgang Pauli and Carl Gustav Jung (see Atmanspacher, 2012). This approach on mental and material problem is strictly associated with ideas and reflections that emerged during the development of quantum theory. The *Pauli-Jung conjecture* had the innovative characteristic of approaching the dual nature (mental and material) in terms of complementarity, a concept itself borrowed from psychology.¹

Pauli and Jung shared the idea of an unknowable reality, in which the material and the subjective worlds are two complementary manifestations of reality. In a letter to Rosenfeld (1952), Pauli claimed that both quantum physics and the psychology of the unconscious need a symbolic psychophysical unitary language for the invisible reality, and this need was the goal he aspired (von Meyenn, 1996, p. 593; also cited in Alcaro et al., 2017, p. 7).

Cognitive neuroscience uses neuroimaging' methods to explore brain structures, functions, and their correlations, in which the functional Magnetic Resonance Imaging (fMRI) is a common non-invasive techniques. This manuscript will focus its approach only on this brain imaging method, specifically through the resting state. Although Hameroff and Penrose (2014) call into question a yet-to-be-discovered theory of quantum gravity to understand how the brain is able to work with non-computable functions (in which microtubules should represent the sites of the associated quantum gates), on the other hand Koch and Hepp (2006) claim that neuroscience does not need quantum physics to explain

the material basis of consciousness, understandable within a purely neurobiological framework. Certainly, a neurobiological approach is necessary in the investigation of mindbrain functioning, but a dialogue with psychoanalysis can be helpful for neuroscientists (Kandel, 1999, 2016). In this sense, quantum physics can be an equally desirable partner for psychoanalysis and cognitive neurosciences in the study of the mindbrain system, consciousness, and self.

This manuscript is not proposing thoughts, affects, drives, and desires as *pieces of brain*, offering a new form of phrenology or localizationism. The proposal is not even based on quantum physics or neuroscientific methods to reveal the secret of self, because those approaches alone cannot answer all the questions about inner human complexity. Nevertheless, it would be naïve to ignore that it is no longer time for a psychoanalytic *brainless* approach in the investigation of the mind, as much as it would be anachronistic to accept a *mindless* neuroscientific approach currently rich in algorithms and poor in subjectivity. The temporospatial dynamics of the brain's spontaneous activity shapes individual mental states, the way the subjects experience themselves and others in time and space (Fingelkurts et al., 2010; Northoff, 2011). The predictive brain, the free energy principle, the study of consciousness, the neural networks—with their features of integration and segregation—and the study of brain entropy in neuropsychological functions and dysfunctions, in the past decade have undoubtedly brought new impetus to the dialogue between mind and brain, between psychoanalysis and neuroscience.

Psychodynamics is the other name of psychoanalysis, coined by the Freud's research supervisor at the University of Vienna: Ernst von Brücke. Together with Hermann von Helmholtz and Emil Du Bois-Reymond, von Brücke was one of the founders of the Physical Society in Berlin (Cieri et al., 2021a). von Brücke and von Helmholtz tried to apply thermodynamics to the psyche, treating the human affects and instincts as energy in physical terms. For instance, to describe the "death drive" (*Todestrieb*) (Freud, 1920) used physics, saying that it is a natural direction of the organic matter to go back to its previous inorganic state; this is an "ancestral dissipation" of energy that follows the entropy direction, expressed by the second law of thermodynamics.

Without forgetting the main clinical component of psychoanalysis, the aim of this paper is to show a link between time and self at the level of a specific neural network, retracing the path of the past, to find a *memoire for the future* of psychoanalysis, in its relationship with physics and neuroscience and remembering that one of the reasons why Freud abandoned this dialogue was the lack of materials and methods, we can nowadays instead rely on.

It will be proposed an exploration of our sense of time in a neural network considered crucial for our self: the Default Mode Network (Raichle et al., 2001), also called Default Network (hereafter DN), closely related to our sense of self, or self-consciousness (Qin and Northoff, 2011; Northoff, 2011; Wittmann, 2013; Wittmann et al., 2015; Davey et al., 2016;

¹The idea of complementarity comes from James (1890) who coined the term, imported by Bohr in the realm of physics, originally with the purpose of replacing the term "wave-particle duality," in his "Como Lecture" in 1927 (Bohr, 1928, p. 566).

Northoff, 2016; Cieri and Esposito, 2019), where the individual sense of past and future seems *entangled*. The spontaneous activity within the DN seems associated with remarkable components of human mental life, such as perception of time and self. To support this, the ontological sense of time will be explored, parallel with the growth of DN's features, limiting the exploration to studies whose authors use the resting state approach, through fMRI. It will also rely on an example from a clinical neuropsychological field, from the most common form of dementia: the Alzheimer Disease (AD), in which the past seems to crumble inexorably without leaving room for any future of the subject. In this form of dementia, the subject's self is equally crushed. Finally, potential repercussions for the clinical and therapeutic dimension will be discussed.

HISTORICAL AND THEORETICAL COMMON GROUND BETWEEN QUANTUM PHYSICS AND PSYCHOANALYSIS

Although at a first glance quantum physics can be perceived as far as possible from psychoanalysis, they share some theoretical and symbolic points that might be worth introducing.

A first connection between psychoanalysis and quantum physics is from a theoretical point of view: both disciplines underline the centrality of relationship. Unlike classical physics, in quantum mechanics the value of a variable is given only at its interactions and this value is only relative to the (other) system affected by the interaction. Here “relative” is used in the same sense in which velocity is a property of a system relative to another system in classical mechanics (Laudisa and Rovelli, 2002). In this context, an electron is considered a set of jumps from one interaction to another; therefore it would not have a discrete position, except in relation to something else (Rovelli, 2014). We can find a speculative parallel in psychoanalysis where relationship is a milestone concept, a *Copernican revolution*, both as a vulnerability component for the development of psychological disorders and a *cure* factor through the psychoanalytic relationship. The attachment, as the first form of relationship, builds not only the baby's relationship mode, or the baby's personality, but rather it shapes his identity, the very existence, the individual self. There is a large psychoanalytic and neuroscientific literature on attachment, as a relationship par excellence. Although this point is beyond the scope of this manuscript, it is possible at least to cite studies by important authors to endorse the common quantum and psychoanalytic idea that “reality is only interaction” (Rovelli, 2014, p. 29). Among these authors, Spitz (1945) pointed out the negative effects of maternal deprivation in early syndromes such as hospitalism, in which children in orphanages do not even have a maternal figure (no affective relationship at all), reported psychophysical damage that can lead to death. Also, Harlow et al. (1965) and Hofer (1994) have

described the central role of attachment through the animal model, while Bion (1962), Bowlby (1958; 1969; 1973), Klein (1948), and Winnicott (1955) have used human model and clinical settings.²

From a neuroscientific perspective, relationship among states, neurons and neural networks is a milestone of cognitive neuroscience and the Free Energy Principle (FEP).³ One of the key concepts of the FEP is the Markov blanket (see Kirchhoff et al., 2018 for details and **Figure 1** for a schematic example). This concept supports the idea that the peculiar function of separation between an internal and an external world allows for relationships between different states. It was born as a statistical method used in machine learning, but its application has been extended further and wider to almost every layer of organization in nature (Clark, 2016; Kirchhoff et al., 2018; Ramstead et al., 2018; Cieri et al., 2021a). In FEP, the Markov blanket divides a system into internal, active, sensory, and external, and without this separation (between states) communication (relationships) would be not possible.

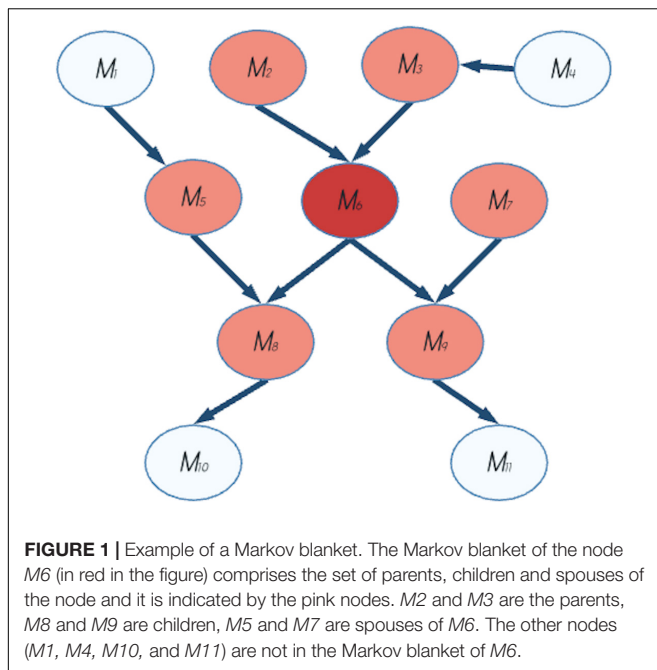
A second aspect in which psychoanalysis and quantum physics can find a communication channel through the lens of psychodynamic neuroscience—the only object of this manuscript—is time. The mystery of time seems to be more about ourselves than about the cosmos; therefore understanding ourselves means reflecting on time, but understanding time means reflecting on ourselves (Wittmann et al., 2015; Rovelli, 2017). In this sense, the current proposal concerns a close neuropsychodynamic relationship between time and self.

Although recognizing the different scales of reality that this manuscript is taking into consideration, comparing psychoanalysis and quantum physics, the attempt will be to find correlations between the human world and the minuteness of the scale at which quantum properties of spacetime (presumably) manifest themselves, recalling a kind of *fractality of reality*, where similar patterns recur at progressively smaller scales. Basically, effects of quantum theory can also appear at larger scales, such as in the case of superconductivity or superfluidity. It is worth mentioning that in 1955 (Pauli, 1955) published a paper on mirror symmetry, expressing a feeling of a deep connections between mind and matter and an inevitable consonance of “inside” and “outside.” In his words:

[...] unconscious motives are always involved thereby. [...] Physics relies on a relation of mirror symmetry between mind and nature (cited in Atmanspacher and Primas, 2006, p. 8).

²In this field, the “experiment” by King Frederick II of Sicily (thirteenth century), who had an interest discovering the natural language of humans, is also sadly known. His question was: if no one speaks with the babies, what language will they learn? The king requested for some children in his kingdom, raising them in total isolation, with no affective relationships, (no attachment), for some years, allowing only specific indicated persons to silently run and leave food for them, with no interactions. Not able to show the “real natural language of human being,” all the children in this “experiment” died from an extreme case stress exposure (Sapolsky, 1994).

³Also, the description of the FEP is beyond the scope of this manuscript. This principle, even in its relationship with psychoanalysis, has been presented in other works (Solms and Friston, 2018; Cieri and Esposito, 2019; Rabeyron, 2021; Solms, 2021).



If, on one hand, general relativity claims that spacetime location is relational only, on the other hand, quantum mechanics tells us that any dynamic entity is subject to Heisenberg's uncertainty at a small scale. Therefore, we need a relational notion of a quantum spacetime to understand Planck scale physics (Rovelli, 1998). The relational nature of reality is studied through quantum mechanics in its appropriate laboratory scale and by psychoanalysis within a clinical setting.

The discussion around the FEP and the neural networks has created the conditions for a new *royal road* for psychodynamic neuroscience (Cieri and Esposito, 2019). In this bridge the concept of entropy plays a key role. Boltzmann's work on entropy can be interpreted through a puzzling conclusion: the difference between past and future can be seen as resulting from a blur to which we are doomed. In other words, entropy exists because we describe the world in a blurry way (Rovelli, 2017).

In *A lover's discourse: fragments* by Barthes and Howard (1978), a child disassembles a watch, to find out what time actually is. Hoping to avoid a similar naivety, in the next sections it will be proposed an attempt to look closer into a specific neural network to see how the past and future are intertwined in our perception of time, and in turn connected with our sense of self. This attempt takes into consideration neuropsychological studies on development, maturation, and impairment of sense of time and self-function through resting state fMRI studies.

THE RESTING STATE

Andreasen et al. (1995) used positron emission tomography (PET) to analyze the neural correlates of random episodic memory, where the uncensored thinking about experience comes to the subject's mind. These authors pointed out that the ability

of the subject to place events in time and to reference them to oneself is at the base of consciousness and self-awareness. They also recalled that the kind of random memory they were investigating is the same used in psychoanalysis through free associations and coined the acronym REST (*Random Episodic Silent Thought*) to indicate this specific form of thinking, now the most used approach in cognitive neuroscience using brain imaging (Cieri et al., 2020). The REST approach comes from the idea to explore the brain's activity when the subject is not involved in any specific cognitive or motor task, lying still in the scanner, a dimension most likely identical to the one identified by the father of American psychology, James (1890), called *stream of consciousness*, 100 years before the REST condition was investigated by brain imaging. This phase was already applied in imaging studies, but it was treated as a control condition to compare the task state. In other words, the investigation of the brain's activation during a task (e.g., a motor task) was the experimental state (or task-positive), which needs a control condition, the resting state, or task-negative, emphasizing its nature partly in contrast with the engagement during the tasks. Both during rest and task, the activity of the brain is organized into intrinsic networks (i.e., brain areas that synchronize their activity) (Raichle et al., 2001). The resting state approach describes how the brain *does not rest, especially during REST*.

As mentioned, the idea that our mental life happens mostly unconsciously is not recent; both Freud and James, from a psychological perspective, claimed that most of our perception is unconscious, emphasizing the importance of the emotional, instinctual and drive processes. They were inspired, and they in turn have inspired philosophers, physicists, and neuroscientists.

On one hand, Schopenhauer's metaphysics had a key role in Freud's theory. On the other hand, Helmholtz's thermodynamics had a fundamental place in building the foundations for the current bridge between neuroscience and psychoanalysis. It is worth mentioning the influence of Schopenhauer (1969) from a philosophical perspective, about the nature and the importance of unconscious processes. Schopenhauer (1969) claimed:

Unconsciousness is the original and natural condition of all things, and therefore is also the basis from which, in particular species of beings, consciousness appears as their highest efflorescence; and for this reason, even then unconsciousness still always predominates (p. 142).

On the physics side, the influence of von Brücke and von Helmholtz on the same topic also played a crucial role in psychoanalysis. When Helmholtz was studying the nervous system, it was a common belief that in the central nervous system, the nerve fibers transmit the electrical energy at the speed of light, therefore our visual perception was reputed instantaneous. In 1852, Helmholtz was able to measure this actual speed, discovering that it took about 20 ms for the nerve impulse to travel 1 m (von Helmholtz, 1971; also cited in Frith, 2007, p. 41). Our perception is driven by involuntary, pre-rational, and reflex-like mechanisms as a crucial part of our formation of visual impressions. Helmholtz called this implicit process *unconscious*

inference, which has had a terrific impact on psychoanalysis, psychology, and the FEP.

The birth and evolution of psychoanalysis have been marked by the dialogue between these authors coming from their respective disciplines. The psychodynamic approach to neuroscience, also thanks to the application of the resting state, is bringing this dialogue back to the center of theory and research, helping the advancement of knowledge toward the consciousness, the unconscious, and the self.⁴

For the main part, the resting state is implicit, involuntary, unconscious, or subconscious. The individual's mindbrain system at REST shows a plenty of activity, and midline brain regions show an increase spontaneous activity (Northoff and Panksepp, 2008), compared to the task-positive state. Performing a specific task increases the brain's energy consumption by less than 5% of the underlying baseline activity. Therefore a large part of the overall activity—from 60 to 80% of all energy used by the brain—occurs in circuits not related to any event of the external world, at the point that different authors (Raichle and Gusnard, 2002; Raichle, 2006; Fox and Raichle, 2007; Zhang and Raichle, 2010) call this intrinsic activity the *brain's dark energy*, a reference to the unseen energy that also represents the mass of most of the universe. This element draws a parallel between the energy in the universe and the energy of mental life, mostly unconscious and unknown, as claimed by Schopenhauer, Freud, James, and von Helmholtz.

DEFAULT NETWORK ASSOCIATED WITH TIME AND SELF

Clinicians have to deal with patients' past, represented by their medical history and patients' future, represented by their prognosis (Lingiardi, 2018). Nonetheless, clinicians know that some neuropsychological diseases are mostly stuck in "one specific time," incapable to perceive different temporal dimensions. For instance, in major depression the patients' past occupies their mind and thoughts, hijacking their individual existence and preventing the development of their future. Minkowski (1995) described this dimension, claiming that the melancholic patients lack the feeling of time as *propulsive energy*, rather feeling blocked by their attention to the past. A similar scenario, in terms of singular dimensional time, is presented by patients affected by post-traumatic stress disorder, in which they have troubles emerging from the past, a traumatic past that somehow suspended their life, not able to see themselves in the future, with memories intruding into the conscious present in form of flashbacks. In these patients is also common to observe physical symptoms, such as fatigue (Di Francesco et al. under review)⁵ underlining the embodied cognition component and its relationship with emotion and time (Wittmann et al., 2015),

symptoms usually accompanied by presence of the timeless dimension of dreams/nightmare.

Another category of individuals with unusual perception of time are the anxious subjects, who show an overestimation of the stimulus durations (Wittmann et al., 2006; Bar-Haim et al., 2010) and report a slowing down of the feeling of time (Lamotte et al., 2014). The psychopathological literature in this field is extensive, including different diagnostic groups and subjective experiences, but the focus of the present manuscript is limited to the "equalization" of past and future at the functional network level of the DN with its possible clinical meanings for the psychodynamic approach. Also, the neuroimaging literature is large on this field, pointing out, among other areas, the key role of the insular cortex, (part of the salience network) in time perception (for meta-analyses, see Lewis and Miall, 2003; Wiener et al., 2010). It is interesting to underline the interactions between insula and the DN, supporting the ability to represent subjects' bodily states to enable conscious reflection on those states (Molnar-Szakacs and Uddin, 2013; Cieri and Esposito, 2019). The anterior insula constitutes a hub involved in the registration of body sensations and filters external salient events, then sending information to the DN that integrates and elaborates information supporting mental activity connected to the self (Seth and Friston, 2016). Although, the present manuscript is focused on the role of the DN it is worth mentioning that the role of insula (with its connection to the DN) points out the concept that "attention to time" is very connected to "attention to bodily signals" (Wittmann et al., 2015).

In our subjective ("healthy") experience we *feel* time, clearly distinguishing past from the future within the DN, memories and inferences (as two opposite ways to think about time) seem *entangled*, meaning a lack of independence between states. During the resting state, healthy adult human being can travel through past, present, and future, in a dynamic temporal dimension described as episodic simulation, which makes it possible to project the self and related events into time (i.e., past and future). The projection into time allows the self (and its related events) to detach or decouple itself from the specific point in time and the current environmental context (Northoff, 2011, p. 126).

The DN seems equally active when the subject makes this time travel, thinking about their past or future (Wittmann, 2013; Andrews-Hanna et al., 2014; Esposito et al., 2017; Cieri and Esposito, 2019); this phenomenon is closely related to the formation and growth of the self during development (Tomasello and Rakoczy, 2003; Fransson et al., 2007; Gao et al., 2009). If time and self-share similar neural correlates in terms of functional connectivity within the DN and between the DN and other networks, the investigation of (statistical) correlation/anticorrelation, studies should show some dysfunction parallel with impairment of sense of time and self.

Cognitive neuroscience exploits fMRI to investigate the variability of the brain through the blood oxygen level-dependent (BOLD) resting state signal (Ogawa et al., 1990) as temporal dynamics of neuronal activity. Functional connectivity (FC) is the temporal correlation (statistical dependencies) between spatially remote neurophysiological events (Friston et al., 1993), while atypical FC changes are observed in several neurological and

⁴Although in psychoanalysis the Self and the Ego are two distinct concepts, where the Ego is one part of the self, this manuscript does not always distinguish between these two instances because from a neuroscientific perspective, their differentiation would be challenging and beyond the scope of this manuscript.

⁵Di Francesco, G., Cieri, F., Sciarra, P., Ballarini, V., Di Ianni, M., and Santarone, S. (under review). Fatigue as mediator factor in PTSD patients after allogeneic hematopoietic stem cell transplantation. *Psychol. Trauma Theory Res. Pract. Policy*

neuropsychiatric disorders, such as mild cognitive impairment, schizophrenia, late life depression, and AD (Buckner, 2013; Andrews-Hanna et al., 2014; Cieri et al., 2017, 2021b, 2022; Esposito et al., 2018; Cera et al., 2019), especially at the level of the DN. Although abnormal functioning of the DN is also clear in individual vulnerability to depression, anxiety, attention deficit, and post-traumatic stress disorders (Slavich and Irwin, 2014), schizophrenia and AD are both specifically considered disconnection syndromes (Geschwind, 1965; Friston and Frith, 1995; Delbeuck et al., 2003; Contreras et al., 2019).

The brain regions that are more active during REST are in the association cortex, including frontal, temporal, and parietal, as well as the retrosplenial cingulate, which shows that during the free associative processes, the association cortices of the human brain communicate to each other easily and without barriers (Andreasen, 2011; Cieri and Esposito, 2018). During the resting state, the neurocognitive awake and healthy system (not affected by disabling neuropsychological disorders) is engaged in the automatic and unconscious registration of bodily, vegetative, endocrine, proprioceptive (internal) inputs and environmental sensorial (external) stimuli, trying to integrate these external and internal stimuli without being overwhelmed and crushed. As mentioned, this activity is mostly bodily, automatic, and unconscious, and, from a temporal perspective it happens in the present, in the *hic et nunc*, the here and now of the subject. During the resting state, the subject is also engaged in a natural free associative dialogue, among emotions, feelings, affects, memories and future plans (Raichle, 2006; Buckner et al., 2008; Northoff, 2011; Christoff et al., 2016; Cieri and Esposito, 2018). The temporospatial dynamics of our brain's spontaneous activity shapes our mental states, the way we experience ourselves and others in time and space (Spagnolo and Northoff, 2021). This spontaneous activity happens in the individual's present, with a strict relationship with past and future. These two temporal dimensions are in fact entangled, intertwined, passing from one thought to another in a plastic and fluid way, experiencing a natural flow of thoughts. Consistently with the hypothesis proposed here and with the FEP (Friston et al., 2006; Friston, 2010) some authors (Buckner et al., 2008) define the DN as a *life simulator*, drawing from past experiences, designing and predicting possible future experiences. The DN plays a key role in the hypothesis testing system (Hohwy, 2013), surfing the uncertainty (Clark, 2016), learning from experience (Bion, 1962) through continuous and uninterrupted unconscious and active inferences (Helmholtz, 1866/1962; Friston, 2010). In other words, this network is close to the establishment, development, and growth of a sense of self, to the point that some authors talk about *Default Self* (Northoff and Bermpohl, 2004; Beer, 2007), suggesting a link between this spontaneous activity and the individual sense of self, self-consciousness (Qin and Northoff, 2011; Wittmann, 2013; Davey et al., 2016; Northoff, 2016) or a dynamic sense of self (Spagnolo and Northoff, 2021).

The brain's topological overlapping between episodic memory and episodic foresight⁶ (Suddendorf, 2006, 2010;

Gilbert and Wilson, 2007) was first noticed by Tulving (1985), who showed that an amnesic subject, with no memories of any episodic event, was similarly impaired in imagining events that they may experience in the future. The episodic memory provides experiential material needed to build episodic foresight, in turn essential to learn from experience. According to Bion (1962), the mental growth and the development of the self depend on the ability of the mind to digest new experiences. We can translate the digestion of new experiences of Bion's approach in the data assimilation or evidence accumulation under the FEP framework (Cieri and Esposito, 2019).

Freud (1900) in *The two principles of mental functioning* claimed that attention is a psychic function established with the aim of exploring the outside world, such that the data received are already familiar when an inner need arises (p. 120). Bion (1962) in *Learning from experience*, pointed out how attention is the alpha function (p. 24), where sensory impressions and emotions are transformed into phenomenon, answering to the characteristics required by the thoughts of the dream. The alteration or the inefficacy of the alpha function causes the beta elements to be unmodifiable, elements seen as the ultimate unknowable truth, chaos, the thing-in-itself, according to the Kantian tradition (p. 27). We need sensory organs as instruments of access to the perception of reality to give sense to it, to transform the non-mental elements, sensory impressions (β -elements), in α -elements, through the α -function to give them an emotional value (Bion, 1962). In fact, when the sensorial systems are impaired (for example as a consequence of a pathological aging process), the cognitive system becomes also impaired, with loss of brain entropy (see Cieri et al., 2021a).

There are many clinical examples of amnesic patients with similar impairment in episodic memory and episodic foresight scenes (e.g., Klein et al., 2002a; Hassabis et al., 2007; Squire et al., 2010; Race et al., 2011). A known patient in this area is also one of the most famous cases in cognitive neuroscience, Henry Gustav Molaison (known as HM) suffered from epilepsy, underwent a bilateral medial temporal lobectomy, surgically resecting the anterior two thirds of his hippocampi, parhippocampal cortices, entorhinal cortices, piriform cortices, and amygdalae. This patient was able to recall everything that happened before his lobectomy, but he was unable to remember or learn memories after his surgery to the point that he shook his neuropsychologist's hand (Brenda Milner), introducing himself with first and last name at any appointment for 40 years, as if it were the first appointment. This patient was equally unable to form episodic memory and episodic foresight. Other works on similar cases confirmed the central role of the hippocampus, underscoring that after bilateral hippocampal damage, patients lose their narrative ability, lacking the power of mental travel either to the past or the future. Their imaginative and planning skills are as strongly impaired as their ability to place themselves into a spatiotemporal context. Instead, they use their logic and semantic knowledge

⁶Episodic foresight is a term proposed by Suddendorf (2006, 2010) to define the ability to imagine future scenarios. Other synonyms are episodic future thinking

(Atance and O'Neill, 2001), envisioning the future (Buckner and Carroll, 2007), episodic simulation of future events (Schacter et al., 2007), and mental time travel into the future (Suddendorf and Corballis, 1997).

to create scenarios that might sound fair (Buzsáki and Llinás, 2017, p. 127).

The episodic memory impairment is the first and most typical symptom in AD and is associated with the similar inability to imagine future scenarios (Addis et al., 2009). In this form of dementia, we observe damages to the hippocampi, parhippocampal, and entorhinal cortices, but also a dysfunction within the DN and between the DN and the other neural networks. In the next paragraph the focus will be the development and decline of the DN, associated with time and self.

DEVELOPMENT AND DECLINE OF THE DEFAULT NETWORK ASSOCIATED WITH TIME AND SELF

“Even a broken clock is right twice a day.”

(The Glass Bead Game, Hesse, 1943).

If episodic memory is correlated to episodic foresight and both are correlated to a sense of self, studies in this field should show a similar ontogenetic increase of memory's capabilities, with parallel ability to imagine future scenario and a contextual growth of the sense of self. That said, typical diseases in which memory degenerates should be correlated with similar impairment of imagining future scenarios, and both should be associated with the impairment of self.

The idea that episodic memory and episodic foresight are correlated, relying on similar neural structures and functions, was proposed by authors (Tulving, 1985; Szpunar, 2010; Schacter et al., 2012; Schacter and Madore, 2016) able to show that remembering the past and imagining future scenarios share similar cognitive and neural processes. In other words, memory is important for connecting the present with the future (Szpunar et al., 2014).

From an ontogenetic perspective, episodic memory formation appears impossible before the age of 4 (e.g., Tulving et al., 2005). Accordingly, before the third or fourth year of age, children usually have a typically confusing idea of positioning events in both past and future. Ability to simulate future scenarios appears in children after the third or fourth year of life (Gilbert and Wilson, 2007), when the brain organizes its spatiotemporal dynamic in the DN meaning that. In other words the brain areas composing the DN are able to synchronize their activity, associated with the ability of anticorrelation with other networks typically active during tasks, such as the mentioned DAN, active during attention tasks. This negative correlation is in turn associated with children's ability to retain and recall memories and create new future scenarios. Consistently, the growth of the self is strictly related with the ability to remember and imagine future scenarios together with the growth of the DN in ontogenetic and phylogenetic senses. This network's connectivity in fact increases through ontogenetic development from birth to adulthood (Fair et al., 2008; Gao et al., 2009). Similarly, the DN regions have undergone significant phylogenetic evolutionary

expansion in modern humans (Van Essen and Dierker, 2007). Here the focus is neither the specific building of episodic memory, nor the neural growth of its specific structures (i.e., hippocampus), but the relationship between past and future, with its connection with a sense of self. From a neural perspective, the focus is on functional connectivity of the DN and its relationship with other neural networks, particularly the DAN.

In a modern healthy adult awake neurocognitive system, the DN and the DAN should show a negative correlation, both during rest and tasks. During rest the DN should be more active, and the DAN should be more silent. During attention tasks the opposite should be true. Consistently, greater DN-DAN negative correlation is associated with higher cognitive control and better working memory performance (Anticevic et al., 2012; Chai et al., 2012).

The self and the DN-DAN anticorrelation grow together during individual development, in conjunction with a more coherent and solid sense of time and self. In fact, parallel with ontogenetic and phylogenetic function and evolution of the self, this anticorrelation appears during the first year of life, strengthens during the second year (Barber et al., 2013), when the child begins to recognize themselves as an object, and becomes stronger in adults to support the development of executive functions and working memory from childhood to adulthood (Andrews-Hanna et al., 2007). The DN-DAN negative correlation seems to show an “inverted-U” shape, increasing during development and decreasing during physiological (Wu et al., 2011), and, especially, cognitively pathological aging (Cieri et al., 2021a). Given this evidence, it is reasonable to observe a decline of this anticorrelation during mild cognitive impairment (Esposito et al., 2018; **Figure 2**) and for it to be even more solid during the AD (Kundu et al., 2019), representing a possible biomarker of neuroaging and cognitive decline.

Late onset AD is the most common cause of dementia; a global health challenge with huge impact in size and costs. In 2021, 6.2 million Americans aged 65 and older living with AD, Alzheimer's Association (2022). Globally, the numbers of people living with dementia will increase from 50 million in 2018 to 152 million in 2050, a 204% increase (Prince et al., 2016). This form of dementia is a complex biopsychosocial disease in which biopsychosocial factors interact in a dynamic and complex way. In this field, there are potential brain imaging biomarkers investigable with different imaging approaches, such as the fMRI.

Although there are different signs, symptoms, and possible biomarkers of AD, one of the first and most disabling neuropsychological signs in these patients is the impairment of episodic memory, where the access to recent events is tragically impaired. Clinicians are more interested in the investigation of the past (memory) of the patients, compared to their planning ability (future), but these temporal dimensions are correlated, lying on the same structure (hippocampus) and the same neural network (the DN). As we have seen, this proposal of a close relationship between time and self at the level of the DN, has shown intriguing parallel among episodic memory, episodic foresight, and a parallel maturation

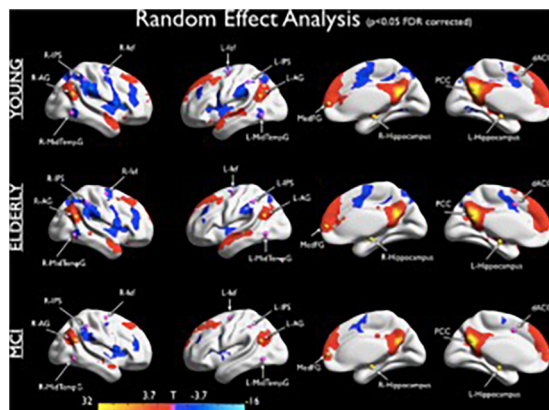


FIGURE 2 | Seed based connectivity maps (seed: PCC) obtained from the random effects group analyses showing the DMN and the anticorrelated DAN for the three groups. The statistical maps were thresholded at $p < 0.05$ (corrected for multiple comparisons using FDR) and superimposed on a partially inflated Talairach template (from Esposito et al., 2018).

of a sense of self. Still, the evidence should show some changes/impairment with pathological aging, when the subject can suffer a loss of cognitive abilities that can begin with episodic memory loss, as in the case of AD. In this case, the resting state-fMRI studies should show some changes at the DN level.

The posterior DN, such as the posterior cingulate cortex and the precuneus, supports autobiographical memory, episodic memory retrieval, future planning, records of bodily sensations, self-reported mental processes, and monitoring psychological states (Borsook et al., 2015; Cieri and Esposito, 2019; Cieri et al., 2021b, 2022). During AD, this brain area is the first affected by atrophy and amyloid- β deposition (Greicius et al., 2003; Buckner et al., 2009). From a functional perspective, we observe a disease progression severity correlated with reduced FC, especially at the level of the posterior DN, as compared to age-matched controls (Zhang et al., 2010; Zhou et al., 2010). According to the hypothesis proposed, the posterior DN has a strategic role in coordinating the interactions among different sensory areas and frames of reference concerning the internal (body) and the external (environment). In fact, bilateral lesions to these areas are related to a virtual breakdown of information integration in the thalamocortical system (Tononi, 2004). AD patients show a similar difficulty in remembering and imagining future scenarios, slipping into a fragmented present without past or future. AD is in fact marked by an inexorable loss of self' functions. In this sense, Proust and Sturrock (2003) words sound extremely appropriate, when he claimed that reality is shaped only in the memory. Despite the concept of self, with its complex functions, is not enclosed or circumscribed in a structure (or a network such as the DN), or in a function (such as memory, see Klein et al., 2002b), in losing their memory, these patients seem to lose their reality, their self.

AD is not the only disorder able to show this phenomenon; other patient populations show episodic memory and episodic foresight impairments, including patients with

temporary amnesia (Juskenaitė et al., 2014), depression (Williams et al., 1996), schizophrenia (D'Argembeau et al., 2008), and post-traumatic stress disorder (Brown et al., 2014).

Lack of memory overlaps with lack of prospect, planning, and imagining future scenarios. This absence of past and future is associated with a loss of reality, therefore a *loss of self*. The function of the DN is related to the self, supporting the internal mental simulations used in adaptive ways (Andrews-Hanna, 2012; Buckner, 2013), consistently with the FEP, in which the system is engaged with simulations, searching of patterns, trying to maintain an internal sensitive balance of the organism, and supporting internal mental simulations used adaptively. In other words, the neurocognitive system is constantly engaged in the search of patterns, trying to build and update its model of the world. In this conscious and unconscious engagement, the DN has a substantial role, mediating between the external and internal stimuli, building dynamic mental future simulations based on past personal experiences. In AD we observe a fall of the individual's time that leaves the patients stuck in a slavery of the present, losing their self' functions.

According to Schacter and Addis (2007):

For more than 100 years, memory has been the object of experimental studies that have focused almost exclusively on its role in preserving and recovering the past. We think it is time to try to understand some of memory's errors by looking to the future (p. 27).

It might be also time to try to understand the relationship between time and self with an approach able to integrate the subjectivity of psychoanalysis with methods of neuroscience, such as the resting state fMRI.

DISCUSSION

"It's a poor sort of memory that only works backwards,"

the Queen remarked.

(Through the Looking Glass; Carroll, 1920).

The DN's processes has been proposed as close to the mediation function attributed by Freud to the Ego, where the system attempts to integrate and elaborate the internal and external stimuli, with the aim of lowering the brain entropy (Carhart-Harris and Friston, 2010; Carhart-Harris et al., 2014; Cieri and Esposito, 2019; Cieri et al., 2021a). The studies described are consistent with the ideas of a proximity between time and self, and from a neuro-functional point of view, perception of time and sense of self share the same neural network: the DN, associated with our *stream of consciousness*. This can have important potential theoretical and clinical significance for psychoanalysis and psychodynamic neuroscience. From a psychodynamic view, the DN might be seen as a sort of neural configuration of the *free floating attention* proposed by Freud (1922), necessary to the psychoanalyst to be in the right "disposition" to listen and feel the stream of consciousness of the patient, allowing the free associative process in both participants of the analytic setting. The free floating attention, together with the free association, are operated by the self, continuously and

constantly built and established over time. Both free floating attention and free associative process are *timeless* state, meaning that they happen in the present, but they are both about past and future, strongly associated to the DN's activity.

Freud (1922) claimed:

[The doctor] should withhold all conscious influences from his capacity to attend and give himself over completely to his “unconscious memory.” The doctor must put himself in a position to make use of everything he is told for the purposes of interpretation and of recognizing the concealed unconscious material without substituting a censorship of his own for the selection that the patient has forgone (p. 111–112).

In 1922, Freud added that the attitude of the clinician should surrender to the unconscious *free floating attention*, also called *evenly suspended attention*, avoiding construction of explicit and conscious expectations (p. 239). A couple of years later (1925), he returned to the topic in his correspondence with Binswanger, specifying that the unconscious must be replaced with the preconscious (Reppen, 2003), a statement that seems to come even closer to the DN's activity. In this scenario, the *free floating attention*, based on the DN, can stimulate the free associative method in both participants—to the point that we might speak about a *default communication*, *default relationship*, or *default therapy* meaning a communication that would seem to involve the DN in a peculiar way, where the history of the individual, their past, present and future narrative, overlap in the same neural process. This can find a symbolic representation in the transferral and controtransferral processes of the psychoanalytic treatment, where the transference can be seen as the unconscious repetition in the here and now of pathogenic conflicts from the past (Kernberg, 2016).

Freud (1933) pointed out the *timeless* of unconscious, writing about the unconscious instance of the id, he claimed that in the id, there is nothing close to the idea of time (p. 106). This idea finds an intriguing overlap with the DN. If we consider the self as composed of the id, the ego, and the superego, the id could intuitively be seen as *the most timeless*. Nevertheless, the self is formed through the biopsychosocial history of the individual, built and place in time.

In the current theoretical framework, both episodic memory and episodic foresight rely on neurocognitive resources at

the DN, suggesting that the link between past and future might have clinical implications, from a neuropsychological and neuroscientific perspective. An important challenge for future research should be the more accurate neurocognitive characterization of episodic memory and episodic foresight, exploring more specifically their common, or different, neural processes. Associating the neuropsychological investigation with a resting state fMRI, would allow a deeper knowledge on neural correlates of both mental processes. This approach can yield important information about our perception of time, our self, and the psychopathological expressions in which these dimensions are impaired, such as AD. Furthermore, the narrative of the psychoanalytic patient could help a process of subjectivity too often missing in the current *mindless* neuroscientific approach. The subjectivity that emerges in the psychodynamic therapeutic relationship could provide useful insights in the study of the perception of time and self.

The importance of *learning from experience* has a tremendous value in terms of human survival, adaptation, and evolution, both as individuals and a species. In this scenario it is crucial to consider the affective environment, the unconscious manifestations, attachment, and relationships with their key role defining the individual's self in time.

Psychoanalysis works with uncertainty and entropy, being a witness and sometimes a guardian (at least of the desire's) chaos. As a patient with an obsessive personality structure, after years of therapy, once wisely said: “I live with the certainty of my uncertainty.” In this patient a neurological condition happened during psychoanalytic psychotherapy, changing in intriguing way the perception of time together with the patient's self.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

FC contributed to the conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing review, and editing.

REFERENCES

- Addis, D. R., Sacchetti, D. C., Ally, B. A., Budson, A. E., and Schacter, D. L. (2009). Episodic simulation of future events is impaired in mild Alzheimer's disease. *Neuropsychologia* 47, 2660–2671. doi: 10.1016/j.neuropsychologia.2009.05.018
- Alcaro, A., Carta, S., and Panksepp, J. (2017). The affective core of the self: a neuro-archetypal perspective on the foundations of human (and animal) subjectivity. *Front Psychol.* 8:1424. doi: 10.3389/fpsyg.2017.01424
- Alzheimer's Association (2022). 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 17, 327–406. doi: 10.1002/alz.12328
- Andreasen, N. C. (2011). A journey into chaos: creativity and the unconscious. *Mens Sana Monogr.* 9, 42–53. doi: 10.4103/0973-1229.77424
- Andreasen, N. C., O'Leary, D. S., Cizadlo, T., Arndt, S., Rezai, K., Watkins, G. L., et al. (1995). Remembering the past: two facets of episodic memory explored with positron emission tomography. *Am. J. Psychiatry* 152, 1576–1585. doi: 10.1176/ajp.152.11.1576
- Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *Neuroscientist* 18, 251–270. doi: 10.1177/1073858411403316
- Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* 1316, 29–52. doi: 10.1111/nyas.12360

- Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., et al. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron* 56, 924–935. doi: 10.1016/j.neuron.2007.10.038
- Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X. J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* 16, 584–592. doi: 10.1016/j.tics.2012.10.008
- Atance, C. M., and O'Neill, D. K. (2001). Episodic future thinking. *Trends Cogn. Sci.* 5, 533–539.
- Atmanspacher, H. (2012). Dual-aspect monism à la Pauli and Jung. *J. Conscious. Stud.* 19, 96–120.
- Atmanspacher, H., and Primas, H. (2006). Pauli's ideas on mind and matter in the context of contemporary science. *J. Conscious. Stud.* 13, 5–50.7.
- Barber, A. D., Caffo, B. S., Pekar, J. J., and Mostofsky, S. H. (2013). Developmental changes in within and between-network connectivity between late childhood and adulthood. *Neuropsychologia* 51, 156–167. doi: 10.1016/j.neuropsychologia.2012.11.011
- Bar-Haim, Y., Kerem, A., Lamy, D., and Zakay, D. (2010). When time slows down: the influence of threat on time perception in anxiety. *Cogn. Emotion* 24, 255–263. doi: 10.1080/02699931.2020.1804329
- Barthes, R., and Howard, R. (1978). *A Lover's Discourse: Fragments*. New York, NY: Hill and Wang.
- Beer, J. S. (2007). The default self: feeling good or being right? *Trends Cogn. Sci.* 11, 187–189. doi: 10.1016/j.tics.2007.02.004
- Bion, W. R. (1962). *Learning From Experience*. London: Heinemann.
- Blass, R. B., and Carmeli, Z. (2007). The case against neuropsychanalysis. On fallacies underlying psychoanalysis' latest scientific trend and its negative impact on psychoanalytic discourse. *Int. J. Psychoanal.* 88, 19–40. doi: 10.1516/6nca-a4ma-mfq7-0tj
- Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature* 121, 580–590. doi: 10.1038/121580a0
- Borsook, D., Maleki, N., and Migraine, B. R. (2015). "Migraine," in *Neurobiology of Brain Disorders*, eds M. J. Zigmond, L. P. Rowland, and J. T. Coyle (Cambridge, MA: Academic Press), 693–708.
- Bowlby, J. (1958). The nature of the child's tie to his mother. *Int. J. Psychoanal.* 39, 350–373.
- Bowlby, J. (1969). *Attachment and Loss, Vol. 1: Attachment. Attachment and Loss*. New York, NY: Basic Books.
- Bowlby, J. (1973). *Attachment and Loss, Vol. 2: Separation: Anxiety and Anger*. New York, NY: Basic Books.
- Brown, A. D., Addis, D. R., Romano, T. A., Marmar, C. R., Bryant, R. A., Hirst, W., et al. (2014). Episodic and semantic components of autobiographical memories and imagined future events in posttraumatic stress disorder. *Memory* 22, 595–604. doi: 10.1080/09658211.2013.807842
- Buckner, R. L. (2013). The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron* 80, 807–815. doi: 10.1016/j.neuron.2013.10.044
- Buckner, R. L., and Carroll, D. C. (2007). Self-projection and the brain. *Trends Cogn. Sci.* 11, 49–57. doi: 10.1016/j.tics.2006.11.004
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The Brain's default network. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., et al. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* 29, 1860–1873. doi: 10.1523/JNEUROSCI.5062-08.2009
- Buzsáki, G., and Llinás, R. (2017). Space and time in the brain. *Science* 358, 482–485. doi: 10.1126/science.aan8869
- Carhart-Harris, R. L., and Friston, K. J. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
- Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., et al. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 8:20. doi: 10.3389/fnhum.2014.00020
- Carroll, L. (1920). *Alice's Adventures in Wonderland*. New York, NY: Macmillan.
- Cera, N., Esposito, R., Cieri, F., and Tartaro, A. (2019). Altered cingulate cortex functional connectivity in normal aging and mild cognitive impairment. *Front. Neurosci.* 13:857. doi: 10.3389/fnins.2019.00857
- Chai, X. J., Castanon, A. N., Ongur, D., and Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *Neuroimage* 59, 1420–1428. doi: 10.1016/j.neuroimage.2011.08.048
- Christoff, K., Irving, Z., Fox, K., Spreng, R. N., and Andrews-Hanna, J. A. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nat. Rev. Neurosci.* 17, 718–731. doi: 10.1038/nrn.2016.113
- Cieri, F. (submitted). Brainless Psychoanalysis: Prejudices and Opportunities in a Difficult Dialogue with Neuroscience.
- Cieri, F., and Esposito, R. (2018). Neuroaging through the lens of the resting state networks. *Biomed. Res. Int.* 2018:5080981. doi: 10.1155/2018/5080981
- Cieri, F., and Esposito, R. (2019). Psychoanalysis and neuroscience: the bridge between mind and brain. *Front. Psychol.* 10:1983. doi: 10.3389/fpsyg.2019.01983
- Cieri, F., Cera, N., Griffo, A., Mantini, D., and Esposito, R. (2020). Editorial: dynamic functioning of resting state networks in physiological and pathological conditions. *Front. Neurosci.* 14:624401. doi: 10.3389/fnins.2020.624401
- Cieri, F., Esposito, R., Cera, N., Pieramico, V., Tartaro, A., and Di Giannantonio, M. (2017). Late-life depression: modifications of brain resting state activity. *J. Geriatr. Psychiatry Neurol.* 30, 140–150. doi: 10.1177/0891988717700509
- Cieri, F., Zhuang, X., Caldwell, J. Z. K., and Cordes, D. (2021a). Brain entropy during aging through a free energy principle approach. *Front. Hum. Neurosci.* 15:647513. doi: 10.3389/fnhum.2021.647513
- Cieri, F., Yang, Z., Cordes, D., Caldwell, J. Z. K., and Alzheimer's Disease Neuroimaging Initiative [ADNI] (2021b). Sex differences of brain functional topography revealed in normal aging and Alzheimer's disease cohort. *J. Alzheimers Dis.* 80, 979–984. doi: 10.3233/JAD-201596
- Cieri, F., Zhuang, X., Cordes, D., Kaplan, N. L., Cummings, J. L., Caldwell, J., et al. (2022). Relationship of sex differences in cortical thickness and memory among cognitively healthy subjects and individuals with mild cognitive impairment and Alzheimer disease. *Alzheimer Res. Ther.* 14:36. doi: 10.1186/s13195-022-00973-1
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and The Embodied Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190217013.001.0001
- Contreras, J. A., Avena-Koenigsberger, A., Risacher, S. L., West, J. D., Tallman, E., McDonald, B. C., et al. (2019). Resting state network modularity along the prodromal late onset Alzheimer's disease continuum. *NeuroImage Clin.* 22:101687. doi: 10.1016/j.nicl.2019.101687
- Craig, A. D. (2009). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 1933–1942. doi: 10.1098/rstb.2009.0008
- Craig, A. D. (2015). *How Do You Feel? An Interoceptive Moment with Your Neurobiological Self*. Princeton: Princeton University Press.
- D'Argembeau, A., Raffard, S., and Van der Linden, M. (2008). Remembering the past and imagining the future in schizophrenia. *J. Abnorm. Psychol.* 117, 247–251. doi: 10.1037/0021-843X.117.1.247
- Davey, C. G., Pujol, J., and Harrison, B. J. (2016). Mapping the self in the brain's default mode network. *Neuroimage* 132, 390–397. doi: 10.1016/j.neuroimage.2016.02.022
- Delbeuck, X., Van der Linden, M., and Collette, F. (2003). Alzheimer's disease as a disconnection syndrome? *Neuropsychol. Rev.* 13, 79–92. doi: 10.1023/a:1023832305702
- Droit-Volet, S., and Gil, S. (2009). The time-emotion paradox. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 1943–1954. doi: 10.1098/rstb.2009.0013
- Esposito, R., Cieri, F., Chiacchiaretta, P., Lauriola, M., Di Giannantonio, M., Tartaro, A., et al. (2017). Modifications in resting state functional anticorrelation between default mode network and dorsal attention network: comparison among young adults, healthy elders and mild cognitive impairment patients. *Brain Imaging Behav.* 12, 127–141. doi: 10.1007/s11682-017-9686-y
- Esposito, R., Cieri, F., di Giannantonio, M., and Tartaro, A. (2018). The role of body image and self-perception in anorexia nervosa: the neuroimaging perspective. *J. Neuropsychol.* 12, 41–52. doi: 10.1111/jnp.12106
- Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., et al. (2008). The maturing architecture of the brain's default network. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4028–4032. doi: 10.1073/pnas.0800376105

- Fingelkurts, A. A., Fingelkurts, A. A., and Neves, C. F. (2010). Natural world physical, brain operational, and mind phenomenal space-time. *Phys. Life Rev.* 7, 195–249. doi: 10.1016/j.plrev.2010.04.001
- Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711. doi: 10.1038/nrn2201
- Fransson, P., Skio cotld, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., et al. (2007). Resting-state networks in the infant brain. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15531–15536.
- Freud, S. (1900). *The Interpretation of Dreams*. London: Hogarth Press.
- Freud, S. (1920). *Beyond The Pleasure Principle*. London: Hogarth Press.
- Freud, S. (1922). *A General Introduction to Psychoanalysis*. New York, NY: Boni and Liveright.
- Freud, S. (1933). “New introductory lectures on psycho-analysis,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. XXII, ed. J. Strachey, trans. (London: Hogarth Press), 1–182.
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. J., and Frith, C. D. (1995). Schizophrenia - a disconnection syndrome. *Clin. Neurosci.* 3, 89–97.
- Friston, K. J., Frith, C. D., Liddle, P. F., and Frackowiak, R. S. (1993). Functional connectivity: the principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* 13, 5–14. doi: 10.1038/jcbfm.1993.4
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle of the brain. *J. Physiol.* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001
- Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. Hoboken, NJ: Blackwell Publishing.
- Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., et al. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6790–6795. doi: 10.1073/pnas.0811221106
- Geschwind, N. (1965). Disconnection syndromes in animals and man. *Brain* 88, 237–294. doi: 10.1093/brain/88.2.237
- Gilbert, D. T., and Wilson, T. D. (2007). Propection: experiencing the future. *Science* 317, 1351–1354. doi: 10.1126/science.1144161
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 100, 253–258. doi: 10.1073/pnas.0135058100
- Hameroff, S., and Penrose, R. (2014). Consciousness in the universe: a review of the ‘Orch OR’ theory. *Phys. Life Rev.* 11, 39–78. doi: 10.1016/j.plrev.2013.08.002
- Harlow, H. F., Dodsworth, R. O., and Harlow, M. K. (1965). Total social isolation in monkeys. *Proc. Natl. Acad. Sci. U. S. A.* 54, 90–97. doi: 10.1073/pnas.54.1.90
- Hassabis, D., Kumaran, D., Vann, D. S., and Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1726–1731. doi: 10.1073/pnas.0610561104
- Helmholtz, H. (1866/1962). “Concerning the perceptions in general,” in *Treatise on Physiological Optics*, 3rd Edn, Vol. III, ed. J. Southall, trans. (New York, NY: Dover).
- Hesse, H. (1943). *Glass Bead Game (Magister Ludi)*. New York, NY: Rinehart and Winston.
- Hofer, M. A. (1994). Hidden regulators in attachment, separation, and loss. *Monogr. Soc. Res. Child Dev.* 59, 192–207. doi: 10.2307/1166146
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001
- James, W. (1890). *The Principles of Psychology*. London: MacMillan.
- Juskenaitė, A., Quinette, P., Desgranges, B., de La Sayette, V., Viader, F., and Eustache, F. (2014). Mental simulations of future scenarios in transient global amnesia. *Neuropsychologia* 63, 1–9. doi: 10.1016/j.neuropsychologia.2014.08.002
- Kandel, E. R. (1999). Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited. *Am. J. Psychiatry* 156, 505–524. doi: 10.1176/ajp.156.4.505
- Kandel, E. R. (2016). *Letà dell'inconscio. Arte, Mente e Cervello dalla Grande Vienna ai Nostri Giorni*. Milano: Cortina Raffaello.
- Kernberg, O. F. (2016). The four basic components of psychoanalytic technique and derived psychoanalytic psychotherapies. *World Psychiatry* 15, 287–288. doi: 10.1002/wps.20368
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface* 15:20170792. doi: 10.1098/rsif.2017.0792
- Klein, M. (1948). A contribution to the theory of anxiety and guilt. *Int. J. Psychoanal.* 29, 114–123.
- Klein, S. B., Rozendal, K., and Cosmides, L. (2002a). A social-cognitive neuroscience analysis of the self. *Soc. Cogn.* 20, 105–135. doi: 10.1521/soco.20.2.105.20991
- Klein, S. B., Loftus, J., and Kihlstrom, J. F. (2002b). Memory and temporal experience: the effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Soc. Cogn.* 20, 353–379. doi: 10.1521/soco.20.5.353.21125
- Koch, C., and Hepp, K. (2006). Quantum mechanics in the brain. *Nature* 440:611. doi: 10.1038/440611a
- Kundu, S., Lukemire, J., Wang, Y., Guo, Y., and Alzheimer's Disease Neuroimaging Initiative [ADNI] (2019). A novel joint brain network analysis using longitudinal Alzheimer's disease data. *Sci. Rep.* 9:19589. doi: 10.1038/s41598-019-55818-z
- Lamotte, M., Chakroun, N., Droit-Volet, S., and Izaute, M. (2014). Metacognitive questionnaire on time: feeling of the passage of time. *Timing Time Percept.* 2, 339–359. doi: 10.1163/22134468-00002031
- Laudisa, F., and Rovelli, C. (2002). “Relational quantum mechanics,” in *The Stanford Encyclopedia of Philosophy (Winter 2021 Edition)*, ed. E. N. Zalta (Washington, DC: The Scholarly Publishing and Academic Resources Coalition).
- Lewis, P. A., and Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Curr. Opin. Neurobiol.* 13, 250–255. doi: 10.1016/s0959-4388(03)00036-9
- Lingiardi, V. (2018). *Diagnosi e destino*. Einaudi, 2018. Collana: Vele.
- Minkowski, E. (1995). *Le temps vécu*. Paris: Presses Universitaires de France.
- Molnar-Szakacs, I., and Uddin, L. Q. (2013). Self-processing and the default mode network: interactions with the mirror neuron system. *Front. Hum. Neurosci.* 7:571. doi: 10.3389/fnhum.2013.00571
- Northoff, G. (2011). *Neuropsychanalysis in Practice: Brain, Self and Objects*. New York, NY: Oxford University Press.
- Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The “basis model of self-specificity” and its encoding by the brain's spontaneous activity. *Cogn. Neurosci.* 7, 203–222. doi: 10.1080/17588928.2015.1111868
- Northoff, G., and Bermpohl, F. (2004). Cortical midline structures and the self. *Trends Cogn. Sci.* 8, 102–107. doi: 10.1016/j.tics.2004.01.004
- Northoff, G., and Panksepp, J. (2008). The trans-species concept of self and the subcortical-cortical midline system. *Trends Cogn. Sci.* 12, 259–264. doi: 10.1016/j.tics.2008.04.007
- Ogawa, S., Lee, T. M., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9868–9872. doi: 10.1016/0005-2736(76)90348-5
- Orwell, G. (1949). *Nineteen Eighty-Four*. New York, NY: Signet Classic.
- Pauli, W. (1955). “Exclusion principle, Lorentz group and reflection of space-time and charge,” in *Niels Bohr and the Development of Physics*, ed. W. Pauli (London: Pergamon Press), 30–51.
- Prince, M., Ali, G. C., Guerchet, M., Prina, A. M., Albanese, E., and Wu, Y. T. (2016). Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimers Res. Ther.* 8:23. doi: 10.1186/s13195-016-0188-8
- Proust, M., and Sturrock, J. (2003). *In Search of Lost Time*. London: Penguin.
- Qin, P., and Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage* 57, 1221–1233. doi: 10.1016/j.neuroimage.2011.05.028
- Rabeyron, T. (2021). Beyond the death drive: entropy and free energy. *Int. J. Psychoanal.* 102, 878–905. doi: 10.1080/00207578.2021.1932514
- Race, E., Keane, M. M., and Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *J. Neurosci.* 31, 10262–10269. doi: 10.1523/JNEUROSCI.1145-11.2011
- Raichle, M. E. (2006). The brain's dark energy. *Science* 314, 1249–1250.
- Raichle, M. E., and Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10237–10239. doi: 10.1073/pnas.172399499

- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., and Gusnard, D. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U. S. A.* 98, 676–682.
- Ramstead, M. J. D., Badcock, P. B., and Friston, K. J. (2018). Answering Schrödinger's question: a free-energy formulation. *Phys. Life Rev.* 24, 1–16. doi: 10.1016/j.plrev.2017.09.001
- Reppen, J. (2003). Ludwig Binswanger and Sigmund Freud: portrait of a friendship. *Psychoanal. Rev.* 90, 281–291. doi: 10.1521/prev.90.3.281.23619
- Rosenfeld, H. (1952). Notes on the psychoanalysis of the super-ego conflict of an acute schizophrenic patient. *Int. J. Psychoanal.* 33, 111–131.
- Rovelli, C. (1998). Loop quantum gravity. *Living Rev. Relativ.* 1:1. doi: 10.12942/lrr-1998-1
- Rovelli, C. (2014). *Sette Brevi Lezioni di Fisica*. New York, NY: Piccola Biblioteca Adelphi.
- Rovelli, C. (2017). *L'ordine del tempo*. New York, NY: Piccola Biblioteca Adelphi.
- Sapolsky, R. M. (1994). *Why Zebras Don't Get Ulcers: A Guide to Stress, Stress Related Diseases, and Coping*. New York, NY: W.H. Freeman.
- Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* 8, 657–661. doi: 10.1038/nrn2213
- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., Szpunar, K. K., et al. (2012). The future of memory: remembering, imagining, and the brain. *Neuron* 76, 677–694. doi: 10.1016/j.neuron.2012.11.001
- Schacter, D. L., and Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. Lond. Ser. B. Biol. Sci.* 362, 773–786. doi: 10.1098/rstb.2007.2087
- Schacter, D. L., and Madore, K. P. (2016). Remembering the past and imagining the future: identifying and enhancing the contribution of episodic memory. *Mem. Stud.* 9, 245–255. doi: 10.1177/1750698016645230
- Schopenhauer, A. (1969). *The World As Will and Representation*. New York, NY: Dover Publication.
- Seth, A. K., and Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philos. Trans. R. Soc. B* 371:20160007. doi: 10.1098/rstb.2016.0007
- Slavich, G. M., and Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol. Bull.* 140, 774–815. doi: 10.1037/a0035302
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. New York, NY: W. W. Norton & Company, Inc.
- Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238.
- Spagnolo, R., and Northoff, G. (2021). *The Dynamic Self in Psychoanalysis: Neuroscientific Foundations and Clinical Cases*. Abingdon-on-Thames: Routledge.
- Spitz, R. (1945). Hospitalism: an inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Stud. Child* 1, 53–74. doi: 10.1080/00797308.1945.11823126
- Squire, L. R., van der Horst, A. S., McDuff, S. G. R., Frascino, J. C., Hopkins, R. O., and Mauldin, K. N. (2010). Role of the hippocampus in remembering the past and imagining the future. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19044–19048. doi: 10.1073/pnas.1014391107
- Suddendorf, T. (2006). Foresight and evolution of the human mind. *Science* 312, 1006–1007. doi: 10.1126/science.1129217
- Suddendorf, T. (2010). Episodic memory versus episodic foresight: similarities and differences. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 99–107. doi: 10.1002/wcs.23
- Suddendorf, T., and Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic Soc. Gen. Psychol. Monogr.* 123, 133–167.
- Szpunar, K. K. (2010). Episodic future thought: an emerging concept. *Perspect. Psychol. Sci.* 5, 142–162. doi: 10.1177/1745691610362350
- Szpunar, K. K., Spreng, R. N., and Schacter, D. L. (2014). A taxonomy of prospection: introducing an organizational framework for future-oriented cognition. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18414–18421. doi: 10.1073/pnas.1417144111
- Tomasello, M., and Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind Lang.* 18, 121–147. doi: 10.1111/1468-0017.00217
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5:42. doi: 10.1186/1471-2202-5-42
- Tulving, E. (1985). Memory and consciousness. *Can. Psychol.* 26, 1–12.
- Tulving, E., Terrace, H. S., and Metcalfe, J. (eds) (2005). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. Oxford: Oxford University Press, 3–56.
- Van Essen, D. C., and Dierker, D. L. (2007). Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron* 56, 209–225. doi: 10.1016/j.neuron.2007.10.015
- von Helmholtz, H. (1971). “The facts of perception,” in *Selected writings of Hermann von Helmholtz*, ed. R. Kahl (Middletown: Wesleyan University Press), 366–408. doi: 10.1093/jrr/rraa135
- von Meyenn, K. (1996). *Wolfgang Pauli. Wissenschaftlicher Briefwechsel, Band IV, Teil I: 1950/1952*. Berlin: Springer.
- Wiener, M., Turkeltaub, P., and Coslett, H. B. (2010). The image of time: a voxel-wise meta-analysis. *Neuroimage* 49, 1728–1740. doi: 10.1016/j.neuroimage.2009.09.064
- Williams, J. M. G., Ellis, N. C., Tyers, C., Healy, H., Rose, G., and Macleod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Memory Cogn.* 24, 116–125. doi: 10.3758/BF03197278
- Winnicott, D. W. (1955). Metapsychological and clinical aspects of regression within the psychoanalytical set-up. *Int. J. Psychoanal.* 36, 16–26.
- Wittmann, M. (2009). The inner sense of time. *Philos. Trans. R. Soc. B* 364, 1955–1967.
- Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nat. Rev. Neurosci.* 14, 217–223. doi: 10.1038/nrn3452
- Wittmann, M., Rudolph, T., Linares Gutierrez, D., and Winkler, I. (2015). Time perspective and emotion regulation as predictors of age-related subjective passage of time. *Int. J. Environ. Res. Public Health* 12, 16027–16042. doi: 10.3390/ijerph121215034
- Wittmann, M., Vollmer, T., Schweiger, C., and Hiddemann, W. (2006). The relation between the experience of time and psychological distress in patients with hematological malignancies. *Palliat. Support. Care* 4, 357–364. doi: 10.1017/s1478951506060469
- Wu, J. T., Wu, H. Z., Yan, C. G., Chen, W. X., Zhang, H. Y., He, Y., et al. (2011). Aging-related changes in the default mode network and its anti-correlated networks: a resting-state fMRI study. *Neurosci. Lett.* 504, 62–67. doi: 10.1016/j.neulet.2011.08.059
- Yovell, Y., Solms, M., and Fotopoulou, A. (2015). The case for neuropsychological analysis: why a dialogue with neuroscience is necessary but not sufficient for psychoanalysis. *Int. J. Psychoanal.* 96, 1515–1553. doi: 10.1111/1745-8315.12332
- Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: MIT press.
- Zhang, D., and Raichle, M. E. (2010). Disease and the Brain's Dark Energy. *Nat. Rev. Neurol.* 6, 15–18. doi: 10.1038/nrneurol.2009.198
- Zhang, H. Y., Wang, S. J., Liu, B., Ma, Z. L., Yang, M., Zhang, Z. J., et al. (2010). Resting brain connectivity: changes during the progress of Alzheimer disease. *Radiology* 256, 598–606. doi: 10.1148/radiol.10.091701
- Zhou, J., Greicius, M. D., Gennatas, E. D., Growdon, M. E., Jang, J. Y., Rabinovici, G. D., et al. (2010). Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain* 133, 1352–1367. doi: 10.1093/brain/awq075

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Cieri. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Daniela Flores Mosri,
Universidad Intercontinental, Mexico
Matthew John Mellor,
University of Exeter, United Kingdom

*CORRESPONDENCE

Thomas Rabeyron
thomas.rabeyron@univ-lorraine.fr

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 27 April 2022

ACCEPTED 19 July 2022

PUBLISHED 09 August 2022

CITATION

Rabeyron T (2022) Psychoanalytic
psychotherapies and the free energy
principle.
Front. Hum. Neurosci. 16:929940.
doi: 10.3389/fnhum.2022.929940

COPYRIGHT

© 2022 Rabeyron. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Psychoanalytic psychotherapies and the free energy principle

Thomas Rabeyron^{1,2,3*}

¹Department of Psychology (Interpsy), University of Lorraine, Nancy, France, ²Department of Psychology (KPU), University of Edinburgh, Edinburgh, United Kingdom, ³Institut Universitaire de France, Paris, France

In this paper I propose a model of the fundamental components of psychoanalytic psychotherapies that I try to explicate with contemporary theories of the Bayesian brain and the Free Energy Principle (FEP). I first show that psychoanalytic therapies require a setting (made up of several envelopes), a particular psychic state and specific processes (transference, free association, dreaming, play, reflexivity and narrativity) in order to induce psychic transformations. I then analyze how these processes of transformations operate and how they can be enlightened by the FEP. I first underline the fact that psychoanalytic therapies imply non-linear processes taking time to unfold and require a setting containing high entropy processes. More precisely, these processes are characterized by an interplay between extension and reduction of free energy. This interplay also favors the emergence of new orders of subjective experience, which occur following states of disorder, according to a certain energetic threshold allowing the modification and improvement of mental functioning. These high entropy states are also characterized by random functioning and psychic malleability which favors the exploration of subjective experience in an original manner. Overall, the approach proposed in this paper support the dialogue between psychoanalysis and other fields of research while underlining how psychoanalytical theoretical and conceptual constructs can also be useful to other disciplines, in particular the neurosciences of subjectivity.

KEYWORDS

psychoanalysis, free energy, entropy, free association, dream, psychoanalytic therapy, transference, symbolization

Psychoanalytic psychotherapies and the free energy principle

In this paper I propose a model of psychoanalytic psychotherapies associated with the free energy principle (FEP). The FEP assumes that any living organism must reduce its internal free energy, which is generated by the tendency of matter toward disorganization due to the second law of thermodynamics (Friston, 2013). This degree of organization, or disorganization, can be measured in physics by the notion of entropy. Living organisms must therefore manage to fight against the increase of entropy (and thus of disorganization) induce by the environment and maintain a sufficient level of organization by reducing free energy. This principle is considered by Karl Friston as being as fundamental as natural selection or gravity (Badcock et al., 2019; Holmes, 2021).

The psychic apparatus can be considered as a hierarchically organized subsystem having the same function of entropy and free energy reduction by developing a set of predictions about the environment based on Bayesian statistics (Friston et al., 2006; Hirsh et al., 2012; Badcock et al., 2019). As summarized by Cieri et al. (2021): “According to the free energy principle, the brain is an open, adaptive, complex system far from equilibrium and as with any adaptive self-organizing biological system in non-equilibrium steady-state with environment, it must reduce its free energy to resist a natural tendency to disorder” (p. 2). Consequently, it might be relevant to try to understand the psychic apparatus from the FEP (Friston, 2013) knowing that, when it is overwhelmed in its ability to regulate free energy, it can rely on a second and external system to regain its internal homeostasis (Friston and Frith, 2015). This process is present from the start of the psychic life and corresponds to complex interrelations between the infant and the attachment figures (Beebe and Lachmann, 2020). Concretely, the states of disorganization of the baby can be regulated by parental care because the psychic apparatus of the baby, as well as its biological functioning in a more general way, is not sufficiently mature to be autonomous. As Winnicott said “there is no such thing as a baby.”

Psychotherapy consists of a relationship whose final aim is to help the patient to find a form of psychic equilibrium and the psychotherapeutic process may also be understood from the point of view of the FEP¹ (Chekroud, 2015; Holmes, 2020, 2021). Psychotherapies correspond to highly complex processes and we still have difficulties in grasping precisely the principles of their action (Cuijpers et al., 2019). They are usually evaluated depending on their theoretical frame of reference (psychoanalytic, cognitive-behavioral therapy, humanistic, systemic, etc.) with the aim of demonstrating their effectiveness for different forms of mental disorders. However, such an approach has led to limited progress, notably because we have great difficulty in discriminating the influence of common and specific factors of therapeutic effectiveness (Cuijpers et al., 2019). In addition, there is often a gap between what a clinician says about what he or she is doing –particularly on the basis of theoretical orientation– and what is actually done in clinical practice (Ablon and Jones, 1998). Consequently, the most important question is not if psychotherapies work, but *how* they work. One of the ways to go beyond these controversies could be to rely on more fundamental principles such as the FEP to better understand the transformations induced by psychotherapies and

thus promote a “meta-perspective of psychotherapy integration” (Holmes, 2020, p. 165). In this regard, the main processes involved in psychoanalytical psychotherapies² will be analyzed in this paper within such a meta-perspective³.

Setting and psychic state during psychoanalytical therapies

I am going first to propose a synthetic model of psychoanalytical psychotherapies identifying several components characteristic of this approach (Figure 1)⁴. I propose to distinguish more precisely: (1) the setting (internal, external and meta), (2) the psychic state of the patient and the therapist, (3) the main processes (transference, associativity, dream, play, reflexivity and narrativity), and (4) specific indicators of the therapy effectiveness (symptoms, object relations, defense mechanisms, anxieties and inhibitions). Each of these components will be described before determining to what extent they can be addressed according to the FEP⁵.

Setting

Psychoanalytic psychotherapies are usually characterized by paying particular attention to the setting (Bleger, 1979). This setting is made up of three complementary envelopes that fit together like Russian dolls: (1) the meta-setting, (2) the external setting, and (3) the internal setting. The meta-setting corresponds to the place where the therapy takes place (for example, a hospital) and more broadly the anthropological and sociological space where the therapy is conducted (for example, in a specific country at a given time). Inside this meta-setting, the external setting corresponds to the space where the therapy takes place (most of the time a consultation

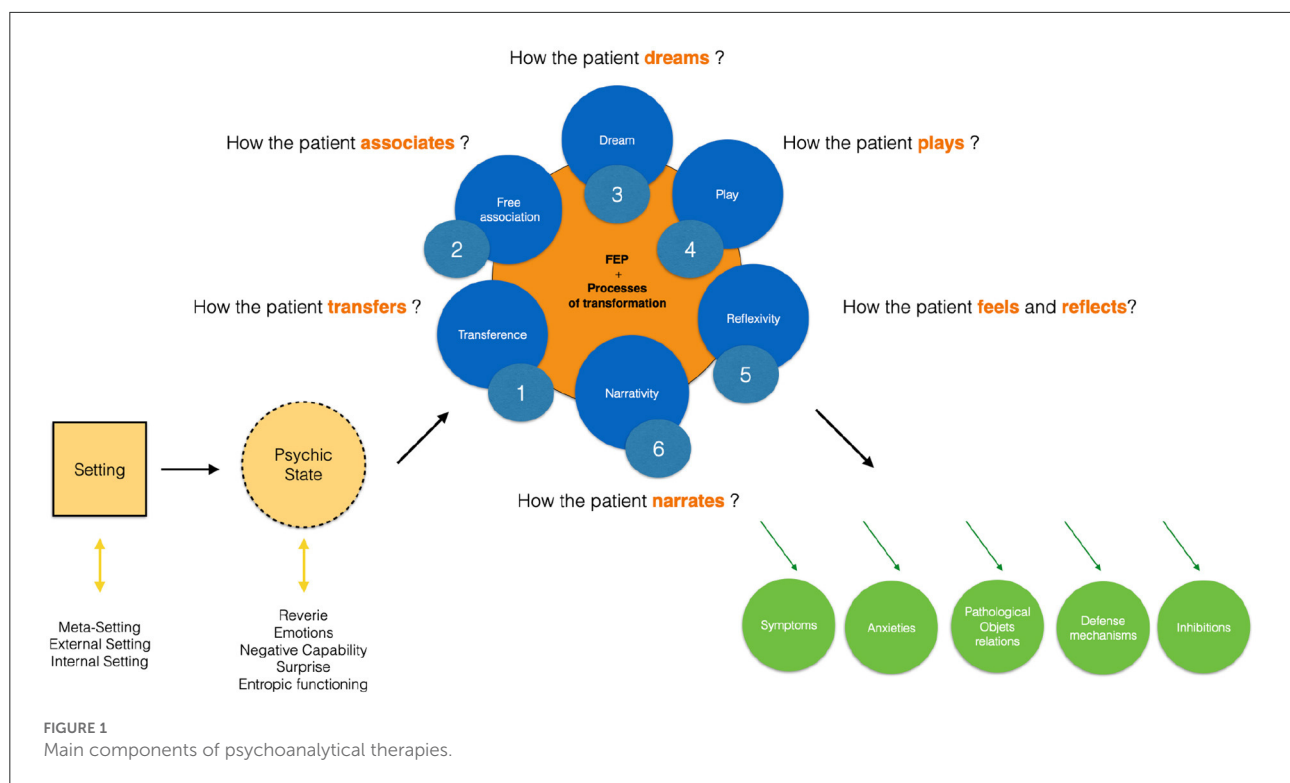
1 The FEP is a “scale free principle”. This characteristic is important because, as Connolly and van Deventer (2017) point out, it means that the free energy principle can potentially be applied to all levels of psychic functioning. It can thus explain purely sensory and perceptual levels of information processing (e.g., visual perception) but also more complex levels of interactions pertaining to the interactions between a therapist and a clinician.

2 whose effectiveness has been largely demonstrated by empirical research (Shedler, 2010; Fonagy, 2015; Leichsenring et al., 2015; Steinert et al., 2017).

3 I have chosen to focus the attention on psychoanalytical psychotherapies because they seem to be particularly compatible with the theories of free energy as already shown in particular by Holmes (2020). But it might be possible to propose a more global, integrative and trans-theoretical model of psychotherapy based on the same approach.

4 I will propose a global approach of psychoanalytical psychotherapies describing mainly what can be considered as common factors of these therapies. I cannot, in this paper, address more precisely the specificities of this approach because there are different schools of psychoanalysis and different categories of mental disorders.

5 This topic would probably deserve a book given its complexity and I will mainly propose in this paper a global overview. The reader looking for a more in-depth analysis of psychotherapy from the point of view of FEP might look at Holmes (2020).



room) and all the elements that characterize this therapeutic space: arrangement of the furniture, decoration, light, noise, etc. These elements create an atmosphere that in itself has an impact on the nature of the processes that will take place during the therapy (Rebollar and Rabeyron, 2016). This setting must also be sufficiently “tough” or “structured” in the sense that it has to allow the patient to feel safe to explore her or his psychic life in complete security. For example, the patient has to be assured that what is said during the sessions is strictly confidential and cannot be overheard outside the consultation room. The third dimension of the setting corresponds to the clinician’s internal setting -also calls “clinical positioning” (Roussillon et al., 2014)- which is based on implicit principles and explicit rules that guide how the clinician interact with the patient and considers the nature of the therapeutic process.

From the point of view of the FEP, the setting appears as a container whose structure allows, as we will see, a double tendency of extension and reduction of free energy during the therapy. In the same manner that the walls of a nuclear reactor contain the reactions induced by nuclear fission, the setting has to be resistant enough to contain the psychological reactions that occur within it⁶. This is all the

more important given that some patients, particularly suffering from borderline personality disorder, regularly seek to “test” the quality of the setting. These patients will also tend to “dialogue” through the setting, i.e., project certain psychological processes onto the setting (Roussillon, 2014). Thus, just as the FEP implies in a living organism the constitution of an envelope to differentiate the internal world from the external world⁷ (Kirchhoff et al., 2018), the setting of a therapy operates as an envelope that contains free energy. Understanding the processes and effects of psychoanalytic psychotherapies therefore implies taking into account the properties of this envelope represented by the different layers of the setting. For example, when this setting is too porous, unstable or even implodes, the therapeutic process will lose its strength, or even collapse, like a cell whose membrane no longer holds

psychic life. This notion is described in this manner in order to insist on the containing and “solid” dimension of the setting, but actually, from a Bionian perspective, the container is a far more labile and porous notion which can also include the patient’s psychic life.

⁷ This envelope can be described, from a mathematical point of view, as a Markov blanket (Kirchhoff et al., 2018). A Markov blanket accounts for certain properties of Bayesian systems, and more precisely for the set of “nodes that surround a node” forming a blanket that surrounds an initial node. This leads to a differentiation between a representation of the internal states of a system and the external states of its environment.

⁶ The notion of container is used in reference to Bion’s writings, insisting here on its “static” configuration, which could give the feeling that the therapist container is entirely differentiated from the patient’s

its role of delimitation between the internal and the external worlds⁸.

Psychic state

Psychoanalytic psychotherapies also rely on the implementation of a specific psychic state shared by the patient and the therapist. This state is favored by free association⁹, which encourages the patient to let go of the usual cognitive psychological functioning (Barratt, 2017; Rabeyron and Massicotte, 2020). The patient is invited to express spontaneously, and without restraint, everything that comes to mind. In this way, the attention is focused on internal experience in a very specific manner leading to the expression of different forms of affects and representations. At the same time, the clinician is also in a specific state of listening called by Freud (1919) “floating attention” which implies freely associating in reaction to the patient’s free associations. The result is a state of shared “reverie” for both the patient and the clinician, which aims to catalyze the processes of symbolization and transformation.

From the perspective of the FEP, this reverie state appears to both promote the increase of emotional and primary processes¹⁰ (which increase free energy) and the decrease of cognitive and secondary processes (which reduce free energy). This psychic state, which is also partly found in certain meditation practices¹¹, also leads to a state of receptivity -that Bion (1965) calls “negative capability”- which allows the patient and the clinician to be surprised by what emerges in the analytical setting (Reik, 1935; Scarfone, 2018). From a Bayesian perspective, surprise is seen as the effect of increased free energy due to a discrepancy between the external world and the a priori predictions created by the brain (Friston et al., 2006). The other forms of emotions and affects are also catalyzed in this psychic state and correspond to an increase of bottom-up processes. In this sense, the setting and a certain “disposition of mind” favors entropic functioning and a form of disorganization of the usual psychological activity during psychotherapy.

⁸ This process can take different forms in the clinic practice: the patient no longer comes to the sessions, the therapist makes a mistake, etc.

⁹ Free association will be analyzed in more details in the third part of the paper.

¹⁰ Especially emotions associated with primary processes as described by Panksepp (1998).

¹¹ With the fundamental difference that meditation can be practiced alone whereas psychoanalytical practices suppose that free association must be done in the presence of another person who is himself in a particular psychic state so that the processes of symbolization operate fully. Another difference is the relationship to time which is more focus on the here and now in meditation while psychoanalysis rely on a larger consideration of time (Cieri, 2022).

Main psychological processes of psychoanalytic psychotherapies and FEP

The setting and the psychic state which characterize psychoanalytic therapies are necessary for the emergence of more specific processes. In this regard, six fundamental processes can be distinguished: transference, free association, dreaming, play, reflexivity, and narrativity. I will briefly outline each of these notions through the prism of the FEP and I will hypothesize that they might help to understand how psychoanalytical psychotherapies lead to a “regulation,” or even a “handling,” of free energy in order to induce psychological transformations. These notions will be presented from the most primary to the most secondary level, knowing, however, that they are deeply interconnected and that distinguishing them in this way is necessarily artificial.

(1) Therapeutic relationship and transference: the therapeutic relationship is like a channel that allows two psyches—those of the patient and the clinician—to become sufficiently attuned to each other to succeed in inducing transformational processes. There is indeed a “we-mode” characterizing the psychic functioning and allowing two people to find, through complex intersubjective exchanges, a new state of equilibrium (Frith, 2012; Friston and Frith, 2015). This we-mode is probably the core of the therapeutic process, as shown by the importance of the therapeutic alliance in therapeutic effectiveness research (Horvath and Luborsky, 1993; Connolly, 2022). However, it is a necessary but not sufficient parameter for more complex processes to emerge. In this respect, the clinician seems to represent an “auxiliary system” for the self of the patient. The therapeutic relationship thus appears to be a privileged tool for reducing free energy insofar as the clinician’s psyche helps to regulate the patient’s psyche, in the same way, metaphorically, that, in Newtonian physics, one object transmits its energy to another object. Therapy can thus be conceptualized, at its most basic level, as an energy transference. For example, a patient in a state of agitation necessarily transfers something of this state to the clinician whose containment allows, at a strictly economic level, to reduce the patient’s free energy as unbound and unelaborated psychic energy. The passage through the other is thus both a vector of regulation and symbolization¹² which implicitly rests

¹² The term symbolization requires definition. The term symbol comes from the ancient Greek *symbolon*, which derives from the verb *symbollesthai*, which means “to put together”. Originally, in ancient Greece, the symbol is more precisely a shard of pottery broken in two, each part of which is given to a contracting party, that testifies the authenticity of the contract, which can be confirmed by putting them together. In a restricted sense, the term symbolization thus designates the operation by which one thing represents another thing.

on a reduction of free energy according to a basic narcissistic identification (Roussillon, 2014). We are then dealing with a very archaic and primary dimension of care relying mainly on the containing, regulating and affective function of the therapist as a mother can be with the infant (Bion, 1962).

This process of intersubjective regulation involves two particular steps. The first is a synchronization which implies that the two psyches “agree” sufficiently to exchange information. This tuning can be analyzed from the point of view of Bayesian systems by taking as a model the synchronization of two birds (Friston and Frith, 2015) and is also supported by the observation of early relationships between babies and their attachment figures (Stern, 1989). From this point of view, psychotherapy can be considered as the meeting of two Bayesian brains which, in order to improve their respective predictions, must agree sufficiently to progressively refine their generative models¹³. This permits a second step of “mixing” of the psyches that goes beyond a simple tuning as noted by Connolly (2022): “the emergence of this synchrony involves a dissolution of boundaries between the two participants where the sensory input from one becomes the prior of the other, as if it emerged from within” (p. 9). Holmes (2020) also explains that “if two inter-acting agents, drawing on the mirror neuron system, mutually assume the other is like themselves, the energetic boundaries between them are temporarily dismantled. Together they build a shared brain or niche” (p. 57). This shared brain evokes what is called the “analytical field” in psychoanalysis, which implies a state of psychic regression for the therapist and the patient in order to produce enduring processes of transformation (Rabeyron, 2020a,b). Such a psychological functioning also seems to catalyzes the transference processes and induces a greater psychic malleability.

These processes are of great complexity and fall under different registers beyond the topic of therapeutic relationship, which leads naturally to the notion of transference and the unconscious dimension of the therapeutic relationship. Transference is classically considered as the way in which the patient spontaneously tends to unconsciously transfer relational patterns onto the figure of the therapist and the therapeutic

relationship (Freud, 1919). This is a central component of psychoanalytic therapies and implies that the clinician accepts to be “taken for another” and be “impregnated” by the patient’s psychic life. For example, the patient will behave toward the therapist in the same way that he or she has behaved toward one of the parents (or, conversely, as one of the parents used to behave toward the patient). These relational patterns are so natural for the patients that they tend to interact with the others, and thus with the therapist, according to the same principles, which can lead to pathological relation patterns taking in particular the form of the repetition compulsion (Freud, 1920). In return, the clinician will be particularly attentive to the countertransference, i.e., the way in which the patient induces, often unconsciously at first, specific relational patterns and ways of thinking that inform the clinician about the patient’s psychic life. How can such a conception of transference can be interpreted from the point of view of the FEP?

First of all, we could consider that the patient express, within the setting of the therapy, relationship patterns which have previously led to a disorganization of the subjective experience. The analysis of transference thus offers the possibility for the patient to play and integrate, in a reflexive manner, specific psychological functioning and pathological relational patterns. This reflexive loop is concomitant with the internalization of therapeutic intersubjectivity which leads to the integration of new styles of emotional regulation. In this sense, transference allows, usually after a long period of therapy, a progressive “updating” of the relational dimension of the generative model, whose primary function is the reduction of free energy. As Freud (1919) already pointed out, the most efficient way to do this “updating” consists in re-playing, in an unconscious way, these relational patterns in the therapeutic setting. The latter then becomes an “attractor” that will favor the emergence of a “transference neurosis” which corresponds to the replay of pathological relational patterns, in the here and now, on the figure of the therapist. This process is what Freud (1919) has called the *agieren* and might be considered, in more contemporary terms, as a process of enaction (Sapichin, 2013).

(2) **Free association** is usually considered the most fundamental rule of analytical practices (Barratt, 2017) and requires a form of “letting go” of mental associations without trying to control them. The patient often goes through several steps and, in particular, through a phase of “focal association” in which an event about reality is reported and from which mental associations will be produced. Gradually, as the therapy progresses, the patient learns -or rather unlearns- to not control his or her thoughts, which leads to a more “unbridled” or “extended” form of free association. This allows, as in certain meditation practices, to observe—with attention and without intention—the internal associative flow of consciousness (Rabeyron and Finkel, 2020). Thus, as explained by Holmes (2020), “The job of free association is to slow things down so they can be identified and subjected to rational top-down scrutiny” (2020, p. 108). This state seems to help the

In psychoanalytical models, symbolization concerns the various stages and processes that allow psychic reality to transform a “primary psychic matter” into a matter endowed with reflexivity and subjectivity. Thus, the symbolization designates how a person transforms his or her own experience in order to integrate it subjectively. We can more precisely distinguish the “primary symbolization” which corresponds to what Freud designates by the term “representation of thing”, while secondary symbolization designates what Freud calls “word representation” and rather concern the processes of psychic transformation associated with verbal language.

¹³ The generative model is, in Friston’s theory, the Bayesian model developed by the brain to predict the propriety of reality and whose main aim is to reduce free energy.

patient to express and explore his internal experience, and to register and translate previous experiences in an original manner, which favors reflexivity and transformational processes (Rabeyron and Massicotte, 2020)¹⁴. The withdrawing from the external world favored by free association results in an increase of free energy and a more chaotic and stochastic experience. Free association then participates in a form of disorganization of the experience which necessarily induces an increase of free energy, which we shall return to in more details later.

(3) **Dreaming and reverie** also appear as fundamental components of psychoanalytical therapies. It concerns both the nocturnal dream activity that the patient may report during the sessions -considering that dreams represent the “royal road to the unconscious” (Freud, 1900)- but it also concerns the state of reverie into which the patient enters during the sessions themselves. In this respect, it is worth remembering that Freud (1940) initially proposed a distinction between free energy and bound energy: “we seem to recognize that nervous or psychic energy occurs in two forms, one freely mobile and another, by comparison, bound” (p. 141). The first corresponds to primary processes in which the energy is considered freer, more mobile, and leads to a rapid discharge in the psychic apparatus. The result is a psychic functioning whose dreaming activity is a typical example and in which a desire might be expressed in a hallucinatory manner. At the opposite, the bound energy corresponds to an energy that is the object of a linking process in the psychic apparatus, notably by cathexis, and whose integration at a higher level of organization induces a more delayed and controlled discharge. There are thus for Freud two levels of “flow” of energy: the first one corresponds to the most primary and unconscious level in which free energy is freer as expressed in the dreams, whereas the second level implies the “handling” of small quantities of energy by consciousness.

Freud (1901) also considers that dreams allow the unconscious processes to be revealed in a more “reliable” manner than in the usual cognitive state as they express more easily unconscious desires and traumatic experiences. Thus, the clinician will pay particular attention to the dreams reported by the patient and may propose interpretations about their meaning. The therapist also plays a specific role during the narrative of the patient’s dream by asking what are the mental associations spontaneously emerging, with the aim to decipher the latent content of the dreams. The dream creates a “virtual reality space” (Hopkins, 2016) that allows the exploration of desires and traumatic experiences that have not been integrated

and that can be expressed in complete safety given that the subject does not act-out in reality the content of the dreams. Then, when the patient becomes capable of elaborating the unelaborated, of making the unconscious conscious, notably through this work focusing on dream expression and analysis, he or she is more in contact with the feelings (internal reality) and with the environment (external reality), and thus refines and perfects the generative model. As proposed by Holmes (2020), the “granulosity” of the world has been improved and more “nuances” are accessible to the patient. As a result, the patient builds a model of reality that diminish free energy because the ability to represent, to predict and to act precisely has been enhanced.

But dream activity cannot be reduced to the nocturnal dreams and there is actually a much bigger “dream function” as shown by Bion (1965). The latter supposes that dream processes are always present in the background of psychological activity, even during the usual cognitive states, and have the function of making perceptive, sensory and emotional experiences “thinkable,” integrable and subjectivable. This process of symbolization through dreams would be catalyzed by the states of reverie that we have already evoked. The patient enters in a state that allows him to dream his experience and to “dream undreamt dreams” (Ogden, 2012), this state being catalyzed by the analyst’s own dream function that Bion (1965) calls “dream-work-alpha.” Dreaming thus appears to be essential in the integration of free energy as Holmes and Nolte (2019) point out: “From the point of view of the FEP, dreaming elaborates the trauma so that it becomes thinkable (p. 9).” This work of psychic integration passes, just as for the free association, by a state of “letting go” which allows the exploration of new states and new configurations of the subjective experience, which, *in fine*, transforms the psychic apparatus and the generative model.

(4) **Play** is also a central component of psychoanalytic therapies which can be considered as “space of play” for both the clinician and the patient in order to explore new experiences. But the patient has to be able to play as Winnicott (1971) pointed out: “Psychotherapy takes place in the overlap of two areas of playing, that of the patient and that of the therapist. Psychotherapy has to do with two people playing together. The corollary of this is that where playing is not possible then the work done by the therapist is directed toward bringing the patient from a state of not being able to play into a state of being able to play” (p. 39). Play appears more precisely as a “state” or a “capacity” that allows unsymbolized and unplayed experiences to be “brought into play” (Roussillon, 2008). But, these the traumatic components of these experiences induce a potential increase of free energy and a disorganization of the psychic apparatus.

Play then implies the regulation of this free energy in a “transitional space” (Winnicott, 1969), i.e., a space of transition between the internal and the external world for experiences not yet integrated. The play is to be distinguished from symptoms

¹⁴ This state is probably associated with specific neurobiological correlates as proposed by Cieri et al. (2021): “during resting state, the brain cortex prominently expresses alpha brainwaves (e.g., 10 Hz), associated with different functions, including top-down inhibitions, top-down prediction and information processing. This activity seems to be associated with the Default Network activity, as part of neural correlated of the ego integrity” (p. 13).

which, on the contrary, are “crystallized” and lead the subject to remain the prisoner of pathological relational patterns and behaviors, which ultimately disturb internal psychological homeostasis. The transitional space also evokes certain cerebral states of “criticality” which corresponds to a transition zone between order and chaos (Cieri et al., 2021): “Critical behavior means that activity of the brain continuously transits between two phases, one in which activity increases and amplifies over time and another in which activity rapidly reduces and dies. Between these two phases, there is a critical zone in which the system increases information processing” (p. 12). The brain alternates between phases of order and disorder, which also characterizes the processes of play and the ability to tolerate a certain degree of disorder to “order” the subjective experience differently. This requires a form of “illusion” corresponding to a particular conjunction between the internal and the external world described by Winnicott (1969) as the “paradox of found-created”: the subject must have the impression that he creates the world where he finds it. This paradox gives the feeling that the representation of the world, as a complex “hallucination,” coincides sufficiently with the world in itself. This is summed up by Frith (2007) “our perception of the world is a fantasy that coincides with reality” (p. 111) and, in terms of the FEP, a sufficient “co-incidence” between the generative model and the environment seems to be necessary to “update” the generative model.

Playing therefore represents an activity allowing the transformation of psychic reality through the exploration of new states of mind. This work of exploration aims in particular at integrating the non-integrated and “playing the non-played,” which underlines the extent to which the *nachträglichkeit* (afterwardness) is a fundamental aspect of psychological functioning. The psychic apparatus is indeed constituted in such a way that there is always a “gap” between what the subject can integrate from his experience (internal and external) and what he perceives or feels of it. In this sense, the play makes it possible to rework non-integrated experiences in the *nachträglichkeit*, and in particular traumatic experiences that could jeopardize the structure of the psychic apparatus or the generative model because of an uncontrolled influx of free energy¹⁵. The play thus

15 This is probably what happens during an episode of decompensation, particularly in the field of psychosis (Ciompi and Tschacher, 2021), which results in the destruction of the “apparatus for thinking the thoughts” (Bion, 1965). The subject is, for example, invaded by the return of traumatic elements in the subjectivity but does not manage to “play” enough with these elements to integrate them. It then produces an implosion of the psychic apparatus and its models of representation leaves a deep trace in the activity of thought from which some subjects never really recover. We can therefore understand the importance for the psyche of regulating these energetic processes because they can result in a deadly danger for the psychic apparatus when the level of free energy reaches a critical level.

appears to be a complex activity of “putting back into play” experiences that have induced too much entropy to be integrated under “good enough” conditions. This is probably the reason why play induces a form of disorder but sufficiently controlled to regulate free energy. Otherwise, the play is “overflowed” and loses its fundamental characteristic of being a play¹⁶.

(5) **Reflexivity** appears as an additional level of psychological functioning which corresponds to the ability to represent the experience according to a “meta” layer (Roussillon, 1991). The human species has developed a degree of reflexivity whose depth is unique in the animal kingdom. This ability allows us to “feel our feelings” and to “think our thoughts,” in an extremely complex manner, which permits us to project ourselves in space and time, thus offering new capacities of adaptation and control of the environment. But this ability has a cost: as the human being becomes more reflexive, he is more distant from his primary perceptive and emotional experience (Roustang, 2003). This results in a “gap” between the more primary layers of experience and the more secundarized and reflective layers of the psychic apparatus. Psychoanalytical psychotherapies are thus based on the idea that the subject may suffer from the non-represented, the non-subjectivated, and that auto-representation processes have potential therapeutic effects. It proceeds to accompany the patient toward an increased reflexivity which allows him or her to feel the not felt and to think the not thought.

We can then differentiate more precisely two main registers of reflexivity: (1) primary reflexivity concerning the relationship to emotions¹⁷; (2) secondary reflexivity which implies linguistic and verbal language. The most primary level of reflexivity concerns essentially how the mental apparatus manages to integrate emotions and their more “entropic” dimension. The second level relates to language and symbolic representations, which organizes the activity of thought presented by Freud as the manipulation of “small quantities of energy” by the psychic apparatus. Reflexivity thus corresponds to a heterogeneous and complex articulation of different levels of psychological functioning¹⁸. The “work of reflexivity” in therapy also implies

16 Play involves a form of “pretending.” It is necessary to “pretend” because this creates a “gap” with the real experience, which encourages its potential integration. When the play “slips”, when we play for “real”, then we no longer play, and the risk is to see the game play overwhelmed in its function of regulation of psychic activity (Rabeyron, 2020a,b).

17 Emotions do not emerge at a reflexive meta layer of mental functioning but rather at very primary layer of psychic life as shown in particularly by Panksepp (1998). Consequently, what is described here as a form of “emotional reflexivity” corresponds to the ability to “feel” emotions –emerging as “bottom-up” processes—that haven’t been “felt” due to, for example, defense mechanisms as cleavage. So, reflexivity deals with the way that emotions will be integrated (and thus felt) at meta layer of mental functioning.

18 Some mental disorders might be the consequence of a disintegration of these registers. For example, certain patients

that the patient speaks and that he uses language to represent his or her internal experience. But this progression through language induces a form of “subjective revolution” and a radical transformation of the relationship between oneself and others (Lacan, 1966). From the point of view of the FEP, a regulation of the free energy occurs by the mediation of language which at the same time represents and regulates the subjective experience in an original manner. As Holmes (2021) underlines “language, whether spoken, sung or gestured, is structured, ordered, negentropic” (p. 2) and language thus induces a new form of order in the psychic apparatus: we pass from an energetic order to a semiotic and signifying order. Through language, the patient “becomes aware” of certain affects and, more generally, how a way of doing things, a dream, a symptom, can be apprehended in an original and “more reflexive” way. Psychoanalytical therapies thus aim at increasing the reflexivity with the idea that this will help to patient be more in phase with the internal and the external worlds, which finally leads to a better regulation of the energetic aspects of the mind necessarily related to the FEP.

(6) **Narrativity** can be considered as a “meta” level of reflexivity. It leads to integrate the reflexive experience into a narrative framework. Reflexivity and narrativity thus maintain a close relationship because reflexivity requires a form of narrativity, an ability to “tell a story” and, conversely, narrativity favors a certain degree of reflexivity. There are different degrees of narrativity of which one of the most primary forms consists of what Bion (1965) has called the “chosen fact” -a term borrowed from the French physicist and philosopher Henri Poincaré- and which corresponds to the linking of several psychic elements (a feeling, a dream, an interpretation, etc.). The chosen fact organizes and generates heuristic emergences that give meaning to the subjective experience. They can be included into narrative scenarios or more complex “narrative envelopes” that Bion calls “vertexes” and that favor the coherence of the experience by creating a more global meaning (Rabeyron, 2018). This narrative framework is fundamental because it permits the subject to inscribe the experience in a story that represents a meta-level of organization of desires, self-esteem, relationships, anxieties, etc. (Mar and Peterson, 2013). More precisely, narrativity implies an origin, a goal, an end, scenarios, protagonists, even a spectator or an “audience,” like in a movie. The human experience would otherwise be reduced to a succession of pictures without coherence and meaning. Thus, some patients suffer from not being able to tell themselves a story, or not being able to tell certain parts of their story, when the trauma has broken into the capacity of subjective elaboration. They can no longer find meaning in their story and psychoanalytic therapies aim to increase, or “repair,” narrativity.

described as “operative” (*opérateur*; Marty, 1998) who seem to think without experiencing emotions, while conversely certain patients are overwhelmed by affects and the inability to symbolize through language expression.

From the FEP perspective, as Holmes (2021) points out “the brain’s aim is constantly to reduce informational entropy and maximize meaning” (p. 2). Narrativity can then be considered as the highest level of subjective experience which nevertheless must be articulated with more basic energetic aspects. In this regard, the brain can be in a “mode” called “fact free learning” or “structured learning” which aims at minimizing complexity by “compressing” representation models (Carhart-Harris and Friston, 2019): “one can refine higher-level models or narratives to make them simpler by removing redundant parameters, thereby revealing the underlying core structures and manifolds” (p. 332). This process leads to “new perspectives resulting in ‘aha’ or ‘eureka’ moments that ‘typically emerge spontaneously, ‘out of the blue’ as simple, elegant solutions, presumably because redundant models and/or model parameters have been unconsciously stripped away, leaving the ‘bare truth beneath’” (p. 332). Thus, the psyche seems to regulate its energetic dynamics by meanings and narratives processes that are supposed to be catalyzed during psychoanalytical therapies.

Transformational processes between chaos and order

Psychoanalytical therapies therefore require a setting (with several envelopes), a specific psychic state and several fundamental processes (transference, free association, dream, play, reflexivity and narrativity) that favor symbolization. These different elements induce effects on the subjectivity, the suffering of the patient and its symptomatic expression¹⁹. In this respect, several “markers” of the therapeutic effectiveness of psychoanalytical therapies can be identified²⁰ (Figure 1): reduction of the symptoms, disappearance of anxieties, decline of toxic object relationships, softening of defense mechanisms and reduction of inhibitions. In comparison to therapies with predetermined goals (focused explicitly on the suppression of symptoms), psychoanalysis assumes, as Freud already proposed, that “healing comes afterwards.” The therapeutic effects will thus become manifest in a second stage, as the “apparatus for thinking the thinking,” as Bion (1965) calls it, will be transformed in depth. This is why it could be preferable not to focus explicitly

¹⁹ An analysis of the relationship between mental disorders and the FEP will not be covered in this paper because the objective is rather to study the underlying secondary processes that can lead to different forms of symptomatic expressions. A more specific analysis of the relationship between psychopathology and FEP can be found in Rabeyron (2021).

²⁰ Other markers could probably be identified (for example, the patient’s autonomy capacities), but it seems that these markers were the most frequently reported in the psychoanalytical literature. It might also be relevant to study in more detail the relationships between these different markers, particularly from the point of view of free energy.

on the symptomatic expression in order to achieve lasting effects on psychic reality and it may be relevant to focus on the source rather than the consequences. From the point of view of the Bayesian brain theories and the FEP, psychoanalytic therapies involve the transformation of the generative model with the idea that this will induce transformation of the subject's relationship to the internal and the external world. This is probably what [Bion \(1965\)](#) has tried to theorize with the “alpha function” which corresponds to the way in which the subject manages to “think” or “dream” the real, thanks to a global “psychic function” which transforms the beta elements (corporal, sensorial, emotional and non-thinkable in themselves) into alpha elements (the first “bricks” of the thought). Psychoanalytic therapies thus aim at transforming the quality of this psychic function of global integration which is necessarily, employed, in its foundations and its most basic logics, toward the need for the minimization of free energy.

The transformation of this global psychic function involves processes of great complexity and non-linear logic ([Galatzer-Levy, 2016](#); [Turtz, 2020](#)). The interconnection of these processes is of such complexity that any attempt to separate them in order to extract empirically testable variables may lead to the destruction of the holistic character of their structure which gives them a “hyper-complex” dimension ([Morin, 2005](#)). We can therefore understand why a therapy probably cannot be carried out in all its depth in just a few sessions. A sufficient frequency and duration are needed for these processes emerge and to induce profound transformations ([Leuzinger-Bohleber et al., 2019](#)). They require time for the transference to be established and for the patient to be able to express freely about intimate experiences. The gradual establishment of a relationship of trust also takes time before allowing the patient and the clinician to be able “play” together and co-create a “narrative” of the therapy, which guarantees the coherence and the meaning of the therapeutic process.

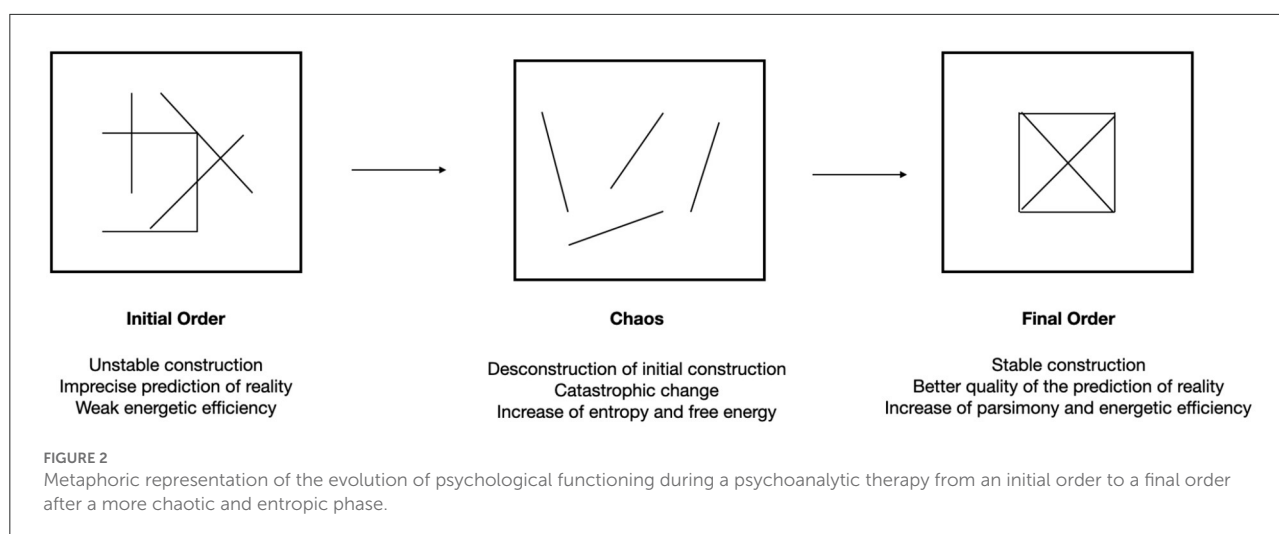
The transformations induced by the therapy can also be envisioned as the consequence of variations of entropy and free energy. From this point of view, let us first recall that the term entropy was initially used by Clausius from the Greek ἐντροπή (ἐντροπή) meaning “contained transformation.” The notion of entropy thus implies, from its origin, the idea of a transformation in a closed system. This evokes several of [Bion \(1970\)](#) notions and in particular this idea: “the analysis in its totality can be seen as a transformation in which an intense catastrophic explosion O has occurred at the affective level” (p. 44). This explosion can be conceived as a high level of free energy that results from a “catastrophic change” contained by the analytic setting. The processes of transformation thus seem to be catalyzed by an increase of free energy within the psychoanalytic setting which allows the containment and the “manipulation” of energy, implicating the setting and the containing dimension of the therapeutic relationship. Thus, for [Holmes and Nolte \(2019\)](#) “the therapeutic duo is an aid to binding potentially

disruptive free energy in creative ways” (p. 6). The therapeutic relationship is thus fundamental to “containing” these entropic logics, otherwise there is a risk that they will overflow or frighten the subject's capacities of elaboration. This implies, as already mentioned, a “negative capability” ([Bion, 1965](#)), on the part of the clinician, which allows the patient to tolerate the uncertainty and the doubt which characterize states of high entropy. Both the patient and the clinician must therefore share a tolerance to uncertainty ([Turtz, 2020](#)) and a “sufficient mental space” for the development of symbolization and maturation processes.

If we try to understand these entropic processes more precisely in the therapeutic setting, they seem to involve a complex interplay between an extension and a reduction of entropy, disorder and order, stochastic and deterministic processes, disinhibition and inhibition. Interestingly, this interplay between order and disorder is also reported in the neuroscience literature as noted by [Cieri et al. \(2021\)](#): “the fact that the brain is considered a system that wanders near a critical dynamic zone between states of order and disorder is widely accepted in cognitive neuroscience” (p. 11). This “double logic” appears to be intimately linked to the processes of transformational processes as noted by [Holmes \(2021\)](#): “a degree of chaos/uncertainty/free energy needs to be tolerated before new generative models can evolve. Homeostatic imprecision needs to be tolerated for a while” (p. 3). The transformation of psychic reality, and the improvement of the generative model, thus implies disorder before creating new forms of order ([Figure 2](#)).

The psychic state that characterize psychoanalytic therapies seems to favor the transition between these different states of organization. The subject can thus explore the internal experience in an original and less organized way than in the usual state of mind, as explained by [Connolly \(2022\)](#): “for stochastic change processes in psychotherapy, Tschacher and Haken have suggested that techniques such as mindfulness, or free association in psychoanalysis, may be examples of interventions that do not push the system state in a particular direction, but may have the effect of making a wider range of experience or behavior possible (p. 4). One of the neurobiological correlates of this state of disorganization and exploration seems to be the Default Mode Network: “It has been proposed that the Default Network has expanded its functions in support of spontaneous cognition. An entropy-extension as one of the processes of human consciousness evolution, subsequently followed by an entropy-reduction might be a specific characteristic adaptable both in phylogenetic and ontogenetic sense” ([Carhart-Harris et al., 2014](#); [Cieri et al., 2021](#), p. 14).

The more entropic states of psychic functioning also appear to be related to an increase of free energy at the level of emotional and primary processes and a decrease of more cognitive and secondary processes. This leads to a state of disorganization, and surprise, which must not, however, limit the subject's integration



capacities, at the risk otherwise of tipping over into fright²¹ as explained by [Holmes \(2020\)](#) “Psychotherapy attempts to create the “duet for one” conditions where surprise becomes allowable and ultimately pleasurable (...) in the form of healing tears” (p. 45). The psychic apparatus must thus combine the need to extract and integrate new information from the internal and external world while accepting the danger of being overwhelmed while processing these new and potential disruptive information. Thus, a balance is needed between these processes of extension and reduction of free energy. We could here take as a metaphor a rice field in which the irrigation of the various plots requires a sufficient supply of water (at the risk otherwise of drying up the crops) but also a regulation of the hydraulic flow (at the risk otherwise of inducing a flood). The psychic apparatus seems to operate according to the same principle with a certain energetic “threshold” that allows to adjust these different processes adequately. Certain mental disorders can then be considered as pathologies of entropic drying ([Holmes, 2020](#)), in particular addictions, anorexia, autism, etc. (e.g., [Wooldridge, 2018](#)). Conversely, other mental disorders can be seen as pathologies of entropic excess: schizophrenia ([Ciompi and Tschacher, 2021](#)), manic episodes, ADHD, etc. ([Rabeyron, 2021](#)).

These high-entropy states can also be approached from the perspective of the randomness they induce in psychic functioning. Free association and daydreaming “soften” associative activity which makes it more random. Like a person who wanders randomly in a forest, the subject then makes discoveries while exploring internal psychic landscapes. He or she is surprised while discovering feelings, images, desires, anxieties, of which he or she did not suspect the existence

and which are then the object of a subjective and reflexive appropriation. This is probably a natural and spontaneous process of the psychic apparatus which is catalyzed by the psychoanalytical setting. Indeed, human beings spontaneously goes through states of reverie as notably shown by the research on mind wandering ([Mooneyham and Schooler, 2013](#)). The psyche needs moments of release from the external world, which induces a state of disengagement from the perceptual, sensorial and motor experience because the psyche cannot, at the same time, transform itself and be in interaction with the world ([Buckner et al., 2008](#)). These transformational processes also imply a certain plasticity of the psyche as illustrated by this metallurgical metaphor proposed by [Connolly \(2022\)](#): “Where chaotic itinerancy is mediated through randomness, its positive effect on the functioning of the system has been formally described as simulated annealing, which is itself analogous to the process of annealing in metallurgy, where a metal is heated up over several occasions, becoming stronger in the process. By the heating process (either real or simulated such as by neural excitability), the system reaches a state of greater plasticity, allowing it to settle again, this time in a new, potentially more optimal attractor state” (p. 5). In the same manner, the psychoanalytical setting “heats” psychic reality and places the patient in a certain state of malleability.

Finally, and in passing, such a state seems to be catalyzed by psychedelics, an observation which may illuminate the nature of these processes. Psychedelics target in particular “high levels of the brain’s functional hierarchy, primarily affecting the precision weighting of high-level priors or beliefs” ([Carhart-Harris and Friston, 2019](#), p. 334). They may induce transformations through “pivotal mental states” which correspond to states of great plasticity²² whose fate

²¹ This state of fright probably corresponds to an emergency procedure which results in the “stopping” of the processes of integration of the environment, at the risk otherwise of “exploding” the generative model as can be seen in psychosis.

²² These states can also be induced by stress as shown also in the field of exceptional experiences ([Rabeyron and Looze, 2015](#)).

can be transformative or pathological²³ (Brouwer and Carhart-Harris, 2021). These states favor in-depth transformation of pathological patterns, thus allowing therapeutic impact on many different clinical symptoms. It could be a common factor with psychoanalytic therapies, given that such states favor the transformation of the structure of the psychic apparatus, which is different from psychotherapies that aim the reduction of specific symptoms. Psychedelics also induce a greater sensitivity to the environment, which also occurs in psychoanalysis when the patient regresses and projects unconscious elements onto the setting (Bleger, 1979). This underlines the fact that the internal malleability requires a malleable environment, as proposed by Winnicott (1969) with the notion transitional space and Milner (1952) with the concept of “pliable medium.” These theories suppose that processes of internal transformation require to be “projected” in the external world, and in particular on objects having specific properties. The psyche has also to be able to “synchronized” sufficiently with the environment to project these psychic contents that are then introjected in order to be symbolized. The activity of symbolization thus requires a passage through exteriority, which means that the internal malleability has to be coupled with an external malleability in order to let the processes of transformation occurring during these states of high entropy.

Conclusion

This paper has proposed a model of the fundamental principles of psychoanalytic therapies articulated with contemporary theories of the Bayesian brain and the FEP. In this view, the main components of psychoanalytic therapies which favor transformational processes have been examined, especially the importance of the setting (with several envelopes), a particular psychic state and specific processes (transference, free association, dream, play, reflexivity, narrativity). The combination of these different elements induces after effects on the patient suffering and

symptoms. Several “markers” of this therapeutic effectiveness on psychic reality have also been identified: reduction of symptoms, disappearance of anxieties, decline of toxic object relationships, softening of defense mechanisms and reduction of inhibitions.

Such a model of psychoanalytical therapies can be compared to some contemporary research in the field of cognitive neurosciences, and in particular those related to the Free Energy Principle (FEP). We can see a certain affinity between psychoanalytical metapsychology and the FEP as pointed out by Holmes (2021): “contemporary psychoanalysis psychotherapy and revitalized Freudian ideas resonate with the FEP” (p. 3). There are indeed several points of similarity between these models which help to construct a global understanding of psychic functioning taking into account the contributions of the science of subjectivity and neurosciences. This also favors the dialogue between psychoanalysis and other fields of knowledge, which puts psychoanalytic metapsychology to the test while underlining how its theoretical constructs could be useful for other disciplines, in particular the neurosciences of subjectivity.

The transformational processes induce by psychoanalytical therapies have also been analyzed by proposing several parallels with the FEP. In this regard, psychological processes seem to implicate non-linear dimensions aiming at transforming the psychic apparatus in its totality. This process of transformation takes time and requires a setting that allows the containment of high entropy processes. More precisely, these processes take place according to an interplay of extension and reduction of free energy. They create new orders from states of disorder according to a certain energetic “threshold” which allow to “update” the generative model. These states of high entropy are also characterized by their random functioning, a necessary preliminary to explore subjective experience in an original manner. States of plasticity associated with this high entropy then favor in-depth transformations of the psyche, which also requires a certain psychic malleability that has to be coupled with a sufficient malleability of the environment. The hypothesis proposed in this work nevertheless remain as a global and conceptual overview and would require much more development to be specific about the relations between psychological processes, psychotherapy and FEP.

Finally, in order to go beyond the thinking proposed in this paper, future developments could attempt to describe more accurately the links between subjectivity, consciousness, information, matter and energy. In this respect, let first recall that matter is condensed energy and that information can be considered as a certain organization of matter that requires energy. From this point of view, psychic reality can be considered as a part of matter with a highly structured organization which requires a considerable quantity of energy to be created, developed and maintained. Psychic reality is, moreover, structured in a heterogeneous way according to a fundamental divide between conscious and unconscious

²³ At a cerebral level, three specific effects of psychedelics are observed (Brouwer and Carhart-Harris, 2021): decreased modularity, decreased hierarchy, and increased overall functionality. This gives the sense that the brain functioning is “flattened” (Carhart-Harris and Friston, 2019) resulting in a “reset” effect. In these very special moments, brain connectivity increases, modularity decreases, and brain plasticity increases. Brouwer and Carhart-Harris (2021) also show specific pharmacodynamics aspects associated with these processes with serotonin 2A receptors that are blocked by psychedelics. The more these receptors are blocked, the greater the subjective effects of the psychedelics will be. The impact on these receptors seems to participate in neuronal plasticity, which would promote long-lasting effects as well as a hyper-sensitivity to the environment when taking psychedelics.

processes. Consciousness thus appears as a “manipulation” of information, notably in the form of representations, and is thus indirectly a form of manipulation of energy²⁴. One could then consider a continuum starting from energy going the most complicated aspects of subjective experience (energy <> matter <> information <> subjectivity). It implies that energetic dimensions that organize matter -in particular the second law of thermodynamics- may have a profound, but indirect and discreet impact, on psychic functioning. What begins to take form in this way could then be a theory that would tackle the hard problem of consciousness (Solms, 2019) from the precise description of such a continuum and its potential discontinuities.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

²⁴ Subjective experience corresponds to a highly complex process constituted by a set of representations whose properties can be studied very precisely thanks to microphenomenology and microanalysis (Rabeyron and Finkel, 2020).

References

- Ablon, J., and Jones, E. (1998). How expert clinicians' prototypes of an ideal treatment correlate with outcome in psychodynamic and cognitive-behavioral therapy. *Psychother. Res.* 8, 71–83. doi: 10.1093/ptr/8.1.71
- Badcock, P. B., Friston, K. J., and Ramstead, M. J. D. (2019). The hierarchically mechanistic mind: a free-energy formulation of the human psyche. *Phys. Life Rev.* 31, 104–121. doi: 10.1016/j.plrev.2018.10.002
- Barratt, B. B. (2017). Opening to the otherwise: the discipline of listening and the necessity of free-association for psychoanalytic praxis. *Int. J. Psychoanal.* 98, 39–53. doi: 10.1111/1745-8315.12563
- Beebe, B., and Lachmann, F. (2020). Infant research and adult treatment revisited: cocreating self-and interactive regulation. *Psychoanal. Psychol.* 37, 313–323. doi: 10.1037/pap0000305
- Bion, W. (1962). *Aux Sources de l'Expérience* (2003^e ed.). Paris: PUF.
- Bion, W. (1965). *Transformations* (2002^e ed.). Paris: PUF.
- Bion, W.-R. (1970). *L'Attention et l'Interprétation* (1990^e ed.). Paris: Payot.
- Bleger, J. (1979). *Psychanalyse du Cadre Psychanalytique. Crise Rupture et déPasement*. Paris: Dunod.
- Brouwer, A., and Carhart-Harris, R. L. (2021). Pivotal mental states. *J. Psychopharmacol.* 35, 319–352. doi: 10.1177/0269881120959637
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Carhart-Harris, R., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., et al. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 8:20. doi: 10.3389/fnhum.2014.00020
- Carhart-Harris, R. L., and Friston, K. J. (2019). REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacol. Rev.* 71, 316–344. doi: 10.1124/pr.118.017160
- Chekrou, A. M. (2015). Unifying treatments for depression: an application of the Free Energy Principle. *Front. Psychol.* 6:153. doi: 10.3389/fpsyg.2015.00153
- Cieri, F. (2022). Memory for the future: psychodynamic approach to time and self through the default network. *Front. Hum. Neurosci.* 16:885315. doi: 10.3389/fnhum.2022.885315
- Cieri, F., Zhuang, X., Caldwell, J. Z., and Cordes, D. (2021). Brain entropy during aging through a free energy principle approach. *Front. Hum. Neurosci.* 15:139. doi: 10.3389/fnhum.2021.647513
- Ciampi, L., and Tschacher, W. (2021). Affect-logic, embodiment, synergetics, and the free energy principle: new approaches to the understanding and treatment of schizophrenia. *Entropy* 23:1619. doi: 10.3390/e23121619
- Connolly, P. (2022). Instability and uncertainty are critical for psychotherapy: how the therapeutic alliance opens us up. *Front. Psychol.* 12:784295. doi: 10.3389/fpsyg.2021.784295
- Cuijpers, P., Reijnders, M., and Huibers, M. (2019). The role of common factors in psychotherapy outcomes. *Annu. Rev. Clin. Psychol.* 15, 207–231. doi: 10.1146/annurev-clinpsy-050718-095424
- Fonagy, P. (2015). The effectiveness of psychodynamic psychotherapies: an update. *World Psychiatry* 14, 137–150. doi: 10.1002/wps.20235
- Freud, S. (1900). *L'Interprétation des Rêves* (2010^e éd.). Paris: PUF.
- Freud, S. (1919). *La Technique Psychanalytique* (1981^e éd.). Paris: PUF.
- Freud, S. (1920). *Au-Delà du Principe de Plaisir* (1979^e éd.). Paris: Payot.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author would like to thank Ian Tierney and the two reviewers for their numerous and particularly useful comments concerning the initial versions of this paper.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Friston, K. (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475
- Friston, K., and Frith, C. (2015). A duet for one. *Conscious. Cogn.* 36, 390–405. doi: 10.1016/j.concog.2014.12.003
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol.* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001
- Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World*. London: Wiley-Blackwell.
- Frith, C. D. (2012). “Implicit metacognition and the we-mode,” in *Workshop on “Pre-reflective and Reflective Processing in Social Interaction”* (Clare College, University of Cambridge).
- Galatzer-Levy, R. M. (2016). The edge of chaos: a nonlinear view of psychoanalytic technique. *Int. J. Psychoanal.* 97, 409–427. doi: 10.1111/1745-8315.12363
- Hirsh, J. B., Mar, R. A., and Peterson, J. B. (2012). Psychological entropy: a framework for understanding uncertainty-related anxiety. *Psychol. Rev.* 119:304. doi: 10.1037/a0026767
- Holmes, J. (2020). *The brain Has a Mind of Its Own: Attachment, Neurobiology, and the New Science of Psychotherapy*. London: Confer Books.
- Holmes, J. (2021). Friston’s free energy principle: new life for psychoanalysis? *Br. J. Psychiatry Bull.* 1–4.
- Holmes, J., and Nolte, T. (2019). “Surprise” and the Bayesian brain: implications for psychotherapy theory and practice. *Front. Psychol.* 10:592. doi: 10.3389/fpsyg.2019.00592
- Hopkins, J. (2016). Free energy and virtual reality in neuroscience and psychoanalysis: a complexity theory of dreaming and mental disorder. *Front. Psychol.* 7:922. doi: 10.3389/fpsyg.2016.00922
- Horvath, A. O., and Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *J. Consult. Clin. Psychol.* 61:561. doi: 10.1037/0022-006X.61.4.561
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., and Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface* 15:20170792. doi: 10.1098/rsif.2017.0792
- Lacan, J. (1966). *Écrits*. Paris: Le seuil.
- Leichsenring, F., Luyten, P., Hilsenroth, M. J., Abbass, A., Barber, J. P., Keefe, J. R., et al. (2015). Psychodynamic therapy meets evidence-based medicine: a systematic review using updated criteria. *Lancet Psychiatry* 2, 648–660. doi: 10.1016/S2215-0366(15)00155-8
- Leuzinger-Bohleber, M., Kaufhold, J., Kallenbach, L., Negele, A., Ernst, M., Keller, W., et al. (2019). How to measure sustained psychic transformations in long-term treatments of chronically depressed patients: symptomatic and structural changes in the LAC Depression Study of the outcome of cognitive-behavioural and psychoanalytic long-term treatments. *Int. J. Psychoanal.* 100, 99–127. doi: 10.1080/00207578.2018.1533377
- Mar, R. A., and Peterson, J. B. (2013). Personal narratives as the highest level of cognitive integration. *Behav. Brain Sci.* 36, 216–217. doi: 10.1017/S0140525X12002269
- Marty, P. (1998). *L'ordre Psychosomatique*. Paris: Payot.
- Milner, M. (1952). Aspects of symbolism in comprehension of the not-self. *Int. J. Psychoanal.* 33, 181–194.
- Mooneyham, B. W., and Schooler, J. W. (2013). The costs and benefits of mind-wandering: a review. *Can. J. Exp. Psychol.* 67, 11–18. doi: 10.1037/a0031569
- Morin, E. (2005). *Introduction to Complex Thinking*. Paris: Le Seuil.
- Ogden, T. H. (2012). *Cet Art qu'est la Psychanalyse : Rêver des Rêves Inrêvés et des Cris Interrompus*. Paris: Ithaque.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.
- Rabeyron, T. (2018). Constructions finies et constructions infinies. *Analysis* 2, 143–155. doi: 10.1016/j.inan.2018.07.002
- Rabeyron, T. (2020a). Le paradigme clinique du jeu. *Bull. Psychol.* 566, 83–98. doi: 10.3917/bupsy.566.0083
- Rabeyron, T. (2020b). Transformational processes and the analytic field: a new paradigm for clinical models and practices. *Psychiatr. Evol.* 85, 101–114. doi: 10.1016/j.evopsy.2019.05.005
- Rabeyron, T. (2021). Beyond the death drive: entropy and free energy. *Int. J. Psychoanal.* 102, 878–905. doi: 10.1080/00207578.2021.1932514
- Rabeyron, T., and Finkel, A. (2020). Consciousness, free energy and cognitive algorithms. *Front. Psychol.* 11:1675. doi: 10.3389/fpsyg.2020.01675
- Rabeyron, T., and Loose, T. (2015). Anomalous experiences, trauma, and symbolization processes at the frontiers between psychoanalysis and cognitive neurosciences. *Front. Psychoanal. Neuropsychol.* 6:1926. doi: 10.3389/fpsyg.2015.01926
- Rabeyron, T., and Massicotte, C. (2020). Entropy, free energy, and symbolization: free association at the intersection of psychoanalysis and neuroscience. *Front. Psychol.* 11:366. doi: 10.3389/fpsyg.2020.00366
- Rebollar, G., and Rabeyron, T. (2016). Clinique et processus de symbolisation de l’atmosphère dans les problématiques psychotiques. *L'Evol. Psychiatr.* 81, 476–492. doi: 10.1016/j.evopsy.2015.05.008
- Reik, T. (1935). *Le Psychologue Surpris: Deviner et Comprendre les Processus Inconscients* (2001e ed.). Paris: Denoël.
- Roussillon, R. (1991). Un sujet qui ne va pas de soi, le sujet en procès. *Rev. Française Psychanal.* 55, 1753–1756. doi: 10.3917/rfp.g1991.55.n5.1753
- Roussillon, R. (2008). *Le Jeu et L'entre-je(u)*. Paris: PUF.
- Roussillon, R. (2014). *Manuel de la Pratique Clinique en Psychologie et Psychopathologie*. Paris: Masson.
- Roustang, F. (2003). *Qu'Est-ce Que L'hypnose?* Paris: Éditions de Minuit.
- Sapisochin, G. (2013). Second thoughts on Agieren: listening to the enacted. *Int. J. Psychoanal.* 94, 967–991. doi: 10.1111/1745-8315.12095
- Scarfone, D. (2018). Free association, surprise, trauma, and transference. *Psychoanal. Inquiry* 38, 468–477. doi: 10.1080/07351690.2018.1480232
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *Am. Psychol.* 65, 98–109. doi: 10.1037/a0018378
- Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714
- Steinert, C., Munder, T., Rabung, S., Hoyer, J., and Leichsenring, F. (2017). Psychodynamic therapy: as efficacious as other empirically supported treatments? A meta-analysis testing equivalence of outcomes. *Am. J. Psychiatry* 174, 943–953. doi: 10.1176/appi.ajp.2017.17010057
- Stern, D. (1989). *Le Monde Interpersonnel du Nourrisson* (2006e ed.). Paris: PUF.
- Turtz, J. S. (2020). Mysteries of the psychoanalytic process: reflections on chaos, complexity, and emergence. *Am. J. Psychoanal.* 80, 176–195. doi: 10.1057/s11231-020-09246-y
- Winnicott, D. W. (1969). *From Pediatrics to Psychoanalysis*. London: Routledge.
- Winnicott, D. W. (1971). *Play and Reality* (1975e ed.). Paris: Gallimard.
- Wooldridge, T. (2018). The entropic body: primitive anxieties and secondary skin formation in anorexia nervosa. *Psychoanal. Dial.* 28, 189–202. doi: 10.1080/10481885.2018.1432949



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

William Sulis,
McMaster University, Canada
Enrico Facco,
University of Padua, Italy

*CORRESPONDENCE

Andrea Scalabrini
andrea.scalabrini@unibg.it

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 28 June 2022

ACCEPTED 03 August 2022

PUBLISHED 02 September 2022

CITATION

Scalabrini A, Mucci C and Northoff G
(2022) The nested hierarchy of self and
its trauma: In search for a synchronic
dynamic and topographical
re-organization.
Front. Hum. Neurosci. 16:980353.
doi: 10.3389/fnhum.2022.980353

COPYRIGHT

© 2022 Scalabrini, Mucci and
Northoff. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The nested hierarchy of self and its trauma: In search for a synchronic dynamic and topographical re-organization

Andrea Scalabrini^{1*†}, Clara Mucci^{1†} and Georg Northoff^{2,3,4†}

¹Department of Human and Social Sciences, University of Bergamo, Bergamo, Italy, ²Faculty of Medicine, Centre for Neural Dynamics, The Royal's Institute of Mental Health Research, Brain and Mind Research Institute, University of Ottawa, Ottawa, ON, Canada, ³Mental Health Centre, Zhejiang University School of Medicine, Hangzhou, China, ⁴Centre for Cognition and Brain Disorders, Hangzhou Normal University, Hangzhou, China

The sense of self has always been a topic of high interest in both psychoanalysis and most recently in neuroscience. Nowadays, there is an agreement in psychoanalysis that the self emerges from the relationship with the other (e.g., the caregiver) in terms of his/her capacity to attune, regulate, and synchronize with the emergent self of the infant. The outcome of this relational/intersubjective synchronization is the development of the sense of self and its regulatory processes both in dynamic psychology and neuroscience. In this work, we propose that synchrony is a fundamental biobehavioral factor in these dialectical processes between self and others which shapes the brain–body–mind system of the individuals, including their sense of self. Recently in neuroscience, it has been proposed by the research group around Northoff that the self is constituted by a brain-based nested hierarchical three-layer structure, including interoceptive, proprio-exteroceptive, and mental layers of self. This may be disrupted, though, when traumatic experiences occur. Following the three levels of trauma theorized by Mucci, we here suggest how different levels of traumatic experiences might have an enduring effect in yielding a trauma-based topographic and dynamic re-organization of the nested model of self featured by dissociation. In conclusion, we propose that different levels and degrees of traumatic experience are related to corresponding disruptions in the topography and dynamic of the brain-based three-layer hierarchical structure of the self.

KEYWORDS

self, synchrony, nestedness, trauma, dissociation, now moments

Introduction

The self then is the sum of everything we are now, and everything we once were, as well as everything we could potentially become

CG Jung, *Mysterium Coniunctionis*, page 108.

The sense of self has always been a topic of interest in both psychoanalysis and more recently in neuroscience.

One of the first psychoanalysts who referred to the concept of self was Carl Gustav Jung, who considered the self as a structure that constitutes the whole of the individual, i.e., more than the sum of its constituents. This gives the self as a whole a conceptually superordinate position in relation to the parts, as it operates autonomously as an overarching organizing principle providing the functional unity of mind and body in a constant flux of states of the organism (refer to Jung, “The dynamic and structure of the psyche,” Volume 8, 2014).

Influenced by Kleinian thought, Fordham suggested how the self is based on Freud’s structural theory. Although “*not explicitly defined... [it] seems to indicate a concept of wholeness which embraces the ego, super-ego, and id, and is even perhaps something more as well*” (Fordham, 1957, p. 198); that is, it implies a wholeness that unites the structural parts. Accordingly, Strachey (1961, pp. 7–8) and Kernberg (1984, pp. 227–8) also noted that Freud himself never dissociated the Ego from the experiencing self; indeed, Freud preserved the German Ich –Ego– as a mental structure and psychic agency, but also as the personal, subjective, experiential self in all his writing. However, differently from the topography proposed by Freud, Jung describes two different complexes that serve the purpose of an interface, on the one hand with the external world, that is the *persona*, and, on the other hand, of an interface with the inner world, that is the *shadow*, to further emphasize the relational feature embedded in its topography of self. In Jungian terms, *persona* is defined as a social relational interface used to underline the prevalence of the world and the others in its constitution; similarly, the shadow can be considered as an interface between the conscious Ego and the inner world. Finally, the self can be considered the center of the psyche to which the other parts are connected and subordinate. Recently, in an interdisciplinary attempt encompassing neuroscience, psychology, philosophy, and anthropology, it has been hypothesized to consider Ego-I-Self as a continuum that reflects different development stages of the sense of individuality, which emerges from the relationship with the external world (Facco et al., 2019). In this hypothesis, the Ego is considered as the primary component appearing in infancy – related to the psychanalytic concept of primary narcissism. The I represents the evolution of the Ego in relation with the principle of reality. The Self includes both the Ego and I and transcends them according with Jung’s individuation process. Although currently, the relational psychodynamic theory has a central role in contemporary psychoanalysis, this perspective was not always recognized. One of the ground-breaking theorists who emphasized the relational component of the development of the self (vs. the intrapsychic conservative component) was Winnicott (1965), who described how the sense of self of the infant can be seen as a result from the

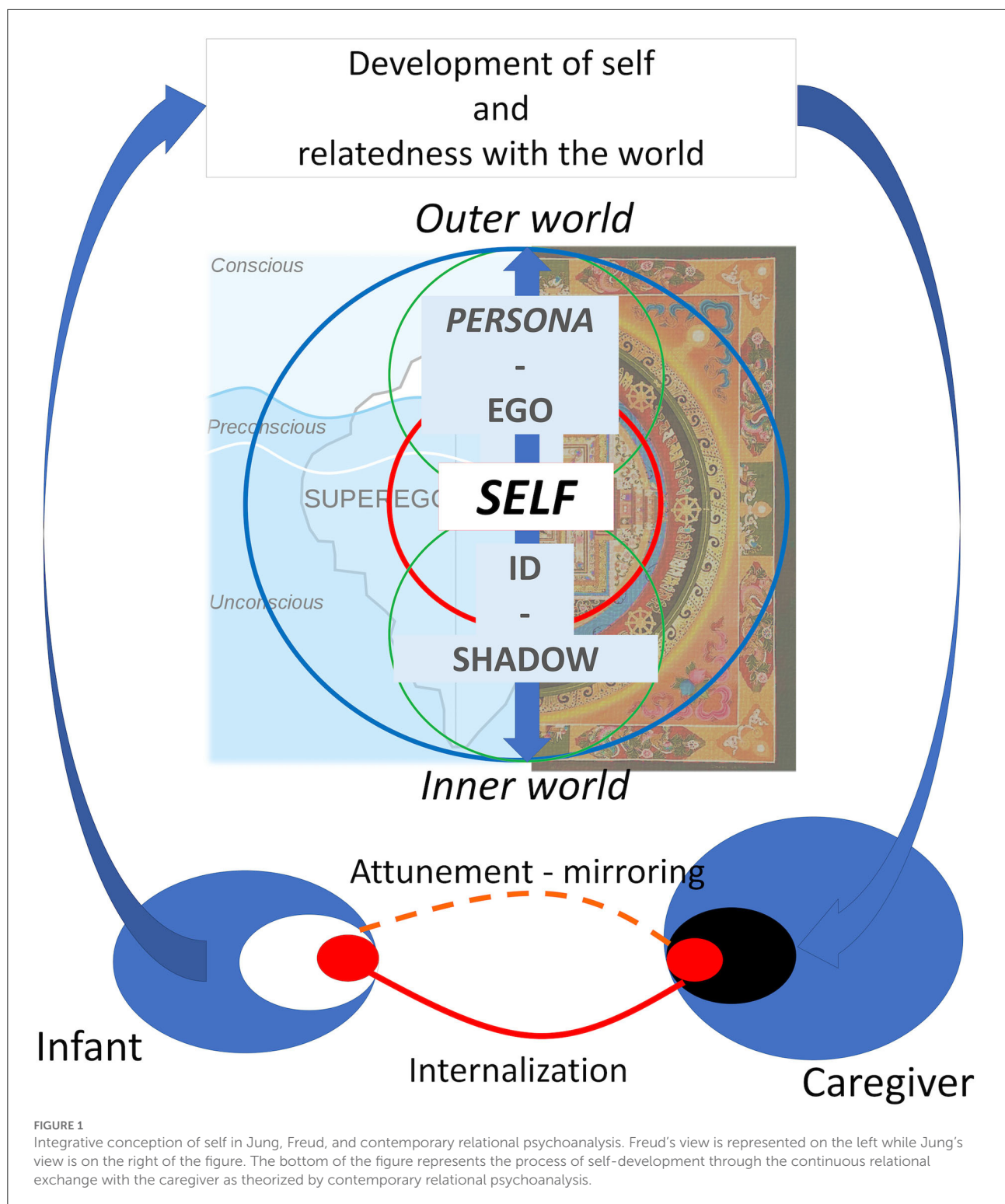
“*internalization*” of the empathic and mirroring relationship with the caregiver.

Similarly, Kohut (1971), who is considered the father of self-psychology, conceptualized how a failure in the development of a cohesive sense of the self, depending on the interaction with the environment, leads to a fragmentation of the body, self, mind, and the self-object. Kohut already pointed out the relevance of the animate environment as fundamental for the development of the sense of self. Indeed, contemporary psychodynamic authors, departing from the background of the attachment theory (Schore, 2000, 2001a,b, 2012; Lyons-Ruth, 2003, 2008; Fonagy et al., 2007; Mucci, 2013, 2018a; Beebe and Lachmann, 2014), proposed that the parent–infant dyad can be considered as the first intersubjective encounter that predisposes the development of the self and emphasized how the dual caregiver–infant exchange continuously modulates the formation of the growing subject, organizing the mind-body-brain interoceptive and exteroceptive connections in relation to the other and the world.

Today, there is an agreement in psychoanalysis that the self comes from the other in terms of his/her capacity to attune, regulate, and synchronize to the emergent self of the infant. The outcome of this relational/intersubjective alignment-attunement-synchronization is the development of the sense of self and its regulatory processes.

Contemporarily, there is a renewed discussion on the relationship between psychoanalysis and neuroscience departing from the “Project for a scientific psychology” (Freud, 1895) reprised recently by Solms in his “New project for a scientific psychology” (Solms, 2020, 2021) and other scholars whom have applied the concept of free energy and predictive coding (Seth and Friston, 2016; Solms and Friston, 2018; Cieri and Esposito, 2019) to explore the mind–brain relationship and the intrinsic relationship between psychoanalysis, the sense of self and its temporal features (refer to Spagnolo and Northoff, 2021; Cieri, 2022).

Our proposal, similar to Pauli and Jung that considered the material and the subjective worlds as two complementary manifestations of reality (refer to Atmanspacher, 2012; Alcaro et al., 2017), aims for searching the “common currency” between the psyche and the brain as shaped by one’s relation with the world (Scalabrini et al., 2020b,c, 2021c; Northoff and Smith, 2022). Here, we propose that synchrony, at a psychological and neuronal level, might be a biobehavioral fundamental factor in the dialectical processes between self and other that shapes the brain–body–mind system of the individual (see Figure 1 for an overview).



The development of self and relatedness through synchrony

Experience of synchrony finds its priors in early relational experiences with primary caregivers in the context of attachment. In 1975, Colwyn Trevarthen proposed his theory about the origin of human intersubjectivity. He posited that the nonverbal communication between mother and infant, *via* the different senses, serves as an intersubjective emotional communication that, through synchronization, regulates the emotional states of both members of the dyad (Trevarthen, 1993). This synchronized protoconversation co-creates an intersubjective reciprocal system of nonverbal communication where “*the emotions constitute a time-space field of intrinsic brain states of mental and behavioral vitality that are signaled for communication to other subjects and that are open to immediate influence from the signals of these others*” (Trevarthen, 1993, p. 155).

Indeed, biobehavioral synchrony is an important aspect of mother–infant attachment (Feldman, 2007), which contributes to the formation of the sense of self and relatedness (Schore, 2005; Mucci, 2018a,b). Through synchronization, the mother regulates the infant’s temperature (Levin, 2006), heart rate (Feldman et al., 2011), sleep, and arousal (Feldman et al., 2002). Mothers regulate their infants’ immune function by breastfeeding, synchronizing their gut microbiota and antigen-specific antibodies (Arrieta et al., 2014). Mothers regulate infants’ arousal with their voice (by singing, or speaking loudly or softly) (Hofer, 1994; Nakata and Trehub, 2004).

Experience of synchrony can be described as the spatiotemporal coordination between the parent’s and the child’s nonverbal behavior and communicative signals during social interactions in ways that enhance positive reciprocity and mutual engagement (Stern, 1985; Trevarthen and Aitken, 2001).

Synchrony can thus be seen as a core mechanism underpinning the development of self and relatedness.

Intriguingly, it has been showed how newborns can discriminate between visuo-tactile synchrony and asynchrony (Filippetti et al., 2013). In this experiment, the infant responded to synchronous or asynchronous tactile stimuli caused by another individual, possibly causing the experience of awareness of his/her own body in the infant. This suggests that the infant might start to rudimentally distinguish between self and others through the experience of synchronization vs. asynchronization. This process has also been called “*mentalization of the body*” (Fotopoulou and Tsakiris, 2017) where the embodied interaction with other people allows the “*mentalization*” of visceral sensation as experienced subjective feelings.

In this regard, a recent natural and ecological study (Ulmer Yaniv et al., 2021) longitudinally tracked how initial mother–child contact increases mother–child synchrony from infancy across the development until young adulthood. The authors investigated how synchronic experiences enhance the

brain’s capacity to empathize with other’s distinct emotions, particularly in the areas of the brain that have been linked with parent–child synchrony in the parental brain, e.g., the insula (Abraham et al., 2016) and the amygdala (Atzil et al., 2011). Data concerning premature neonates showed that the proximity to the mother’s body during incubation is linked with enriched synchronous infant–caregiving experiences reverberating throughout the course of child development. The researchers suggested that synchrony is a natural mechanism by which the human brain connects to others and the social world. In this context “*attachment begins before any sense of self and before any sense of object to attach to*” (Brockman, 2002, p. 90). Indeed, the mother–infant attachment relationship is considered a prior to the development of social synchrony, constituting a dynamic repertoire that can be considered as the capacity to automatically integrate others’ emotions together with the abilities to use interoceptive signals to detect others’ specific effect as distinguished from one’s own effect (self–other distinction).

It has also been shown how the synchronization capacity is affected by the initial opportunities of physical contact in the context of mother–infant attachment. Consistent with the dynamics system theory, preterm infants lacking opportunities of a full maternal physical contact consistently showed lower levels of synchrony compared to infants that did receive contact (Feldman, 2015). Finally, to investigate empathic abilities, the same subjects at the age of 20 performed an fMRI empathy task that involved affective-empathic network constituted prominently by ventromedial prefrontal cortex, temporal pole, insula, and amygdala. The involvement of these regions corresponded to the degree of emotional intensity and the enhanced affective-specific neuronal response in amygdala and insula was related to the synchrony experienced from infancy to adulthood. Focusing now on mothers, it has also been showed that maternal insula and amygdala responses to affective stimuli related to an infant (e.g., infant’s crying and laughing) also predict mother–infant synchrony, oxytocin receptors, and play a causal role in the initiation of mammalian mothering. On the other hand, when there is a lesion of the amygdala, maternal caregiving is disrupted (Riem et al., 2011; Lonstein et al., 2015).

Similarly, the human parents’ insular response to their infant impacts the child’s later capacity to regulate distinct emotions, both positive emotions and distressful ones, in preschool (Abraham et al., 2016). Given that the link between the parental insula and the child’s regulation of distinct emotions was again mediated by parent–infant synchrony, it seems that regions of the brain, such as the amygdala and the insula, may provide an integrative foundation for the cross-generational transmission of parenting as mediated by synchronous caregiving.

These findings relate to biobehavioral synchrony, *via* the consistency of caregiving over time, which seems to enhance the development of the child’s sense of self and relatedness.

These first synchronous experiences, facilitated through the insula and amygdala as a part of the affective-limbic system, seem to shape the development of a physiological and psychological “baseline” (refer to Northoff, 2016; Scalabrini et al., 2021c, 2022; Northoff et al., 2022) that later assemble predictive models that facilitate the development of a social brain and the self (Apps and Tsakiris, 2014; Atzil et al., 2018) (see also Ebisch et al., 2022 for the role of the insula and empathic traits). The baseline means that the self serves as a reference or standard for any subsequent input processing; more importantly, the self as the baseline serves as a reference for subsequent cognition, affect, and other functions – this is well expressed in the concept of the “psychological baseline” (Northoff and Bermpohl, 2004; Scalabrini et al., 2021c, 2022) and the baseline model of self-specificity (BMSS; Northoff, 2016; Northoff et al., 2022; Scalabrini et al., 2022).

The role of the self as the baseline for any subsequent psychological function is constituted by the above-described dialectical process through synchronization as “relational alignment” (Schore, 2012; Scalabrini et al., 2018; Mucci, 2021a,b) that facilitates the development of the integration of interoceptive signals and exteroceptive information. This is essential for the child’s homeostasis and for the development of a social-cognitive-affective-behavioral repertoire related to the sense of self (Craig, 2003; Fotopoulou and Tsakiris, 2017; Atzil et al., 2018).

The quality and the organization of this psycho-physiological baseline needs to be considered as dependent on the relational experiences, first with the attachment figure (e.g., the mother) and, second, with an extended social world which can impact the self both in a positive and in an adverse/traumatic way (refer to Mucci and Scalabrini, 2021). Intriguingly, such internalization of subjective and intersubjective experiences can be rooted in the spontaneous spatiotemporal neuronal activity “baseline” of the brain (Northoff et al., 2020a,b, 2022; Northoff and Scalabrini, 2021). Such baseline on the neuronal level of the brain may well be related to the above-described baseline on the more psychological level, the psycho-physiological baseline: we propose that both baselines are linked and connected through their shared spatiotemporal features, e.g., topography and dynamic as their “common currency” (Northoff et al., 2020a,b).

Neuronal synchronization: Insula at the crossroads between brain–body–mind

Synchronization at a neuronal level plays an essential role in building the complex spatiotemporal structure and dynamics of the brain. Neuronal synchronization allows integrating neuronal activity from different brain regions (and their respective psychological functions and contents) over long stretches of time and distant regions/networks. Since the correlation between the

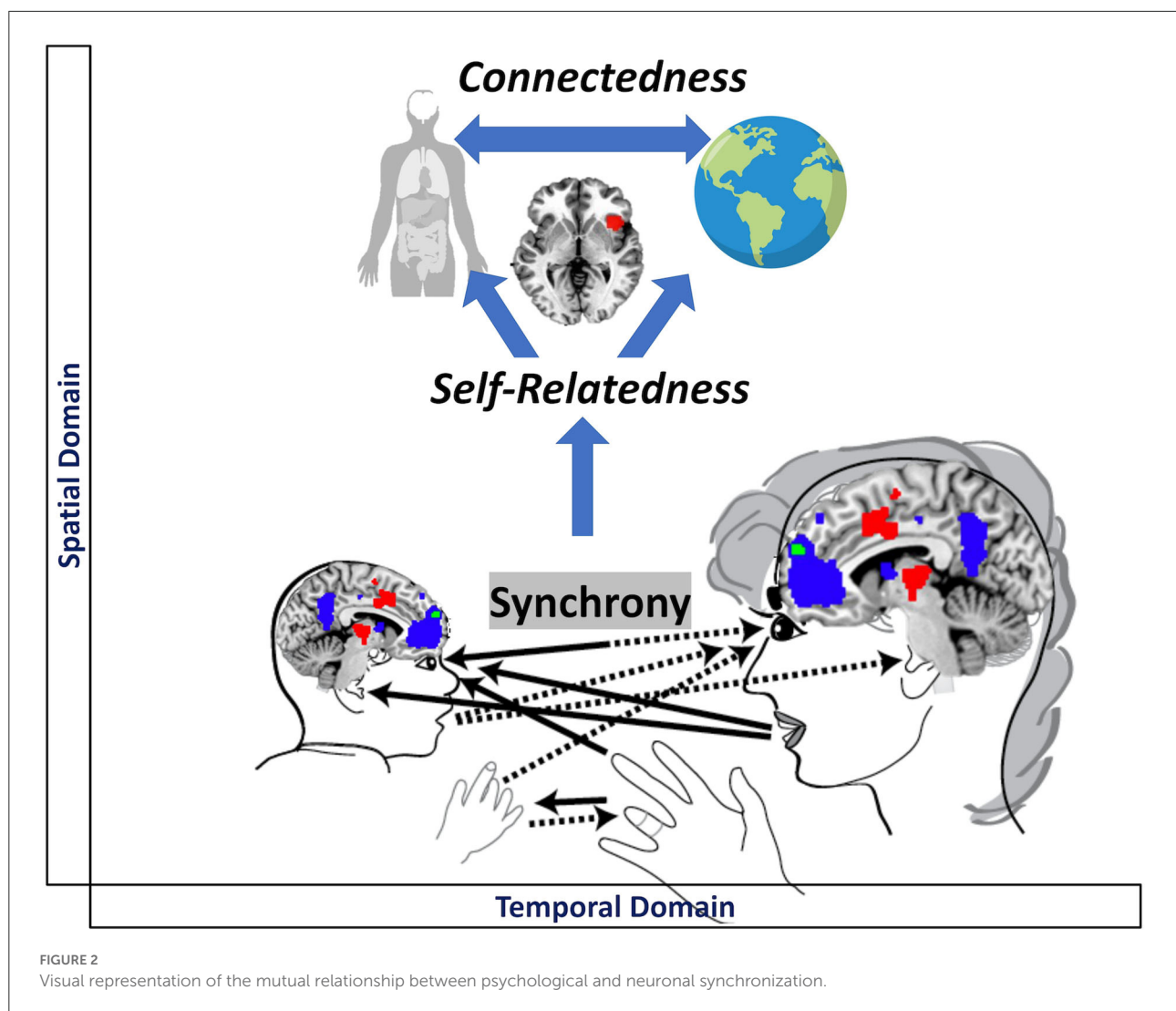
time series entails synchronization between different regions’ neuronal activity, one might suppose that synchronization plays a pivotal role also at a psychological–phenomenological level.

For instance, the group around Tallon-Baudry demonstrated how our psychological sense of self is based on body–brain coupling by temporal synchronization of interoceptive stimuli from the heart and stomach with the brain’s spontaneous activity in the insula and other regions, such as the anterior midline regions and visual cortex (Park and Tallon-Baudry, 2014; Park et al., 2014). These findings show that the brain and the temporal structure of its neural activity align or synchronize themselves to the ongoing temporal structure of the body and its ongoing visceral activity in the stomach and heart (Northoff and Huang, 2017; Northoff, 2018). Such “temporo-spatial alignment” (Northoff and Huang, 2017; Northoff, 2018) of the brain to the body, e.g., neuro-visceral monitoring (Tallon-Baudry et al., 2018), is central for constituting the self (first-person perspective). Thus, our sense of self, e.g., our subjective first-person perspective, is based on how our brain is aligned to and thus synchronized with the body, and ultimately the environment or world (Northoff, 2018). Intriguingly, it has been shown at a psychological level that the higher the degree of perception of synchronicity in time and closeness in space toward another person (the animate target) vs. an object (the inanimate target) the higher the degree of self-relatedness (Scalabrini et al., 2019). Similarly, studies conducted by the group around Ansermet and Magistretti investigated the subjective sensation of synchrony or connectedness during a joint task suggesting that the more similar the movements were, the sooner the sensation of synchrony appeared (Llobera et al., 2016).

The construct of self-relatedness here seems to be very close to the larger construct of connectedness that has been used to describe the connection to the body (Porges, 2011), the self, to others, and the world in general (Watts et al., 2017). Congruently, in the field of psychedelic science, a group of scholars (Carhart-Harris et al., 2018) aim to develop an operational definition of connectedness that incorporates not just connectedness in the subjective sense, but also at its biological and behavioral levels. Recently in dynamic psychology and psychoanalysis, Mucci (2013, 2018a, 2022) proposes the construct of connectedness as fundamental in the social exchanges of human beings and in maintaining a bond to life and is itself a form of resilience against adverse life traumatic experiences that might disrupt our sense of self and relatedness with others (refer to Figure 2).

The multilayer nested model of self

The attention of several authors has recently been elucidating the neural architecture related to the self. For example, Panksepp (1998) and Damasio (2010) associated

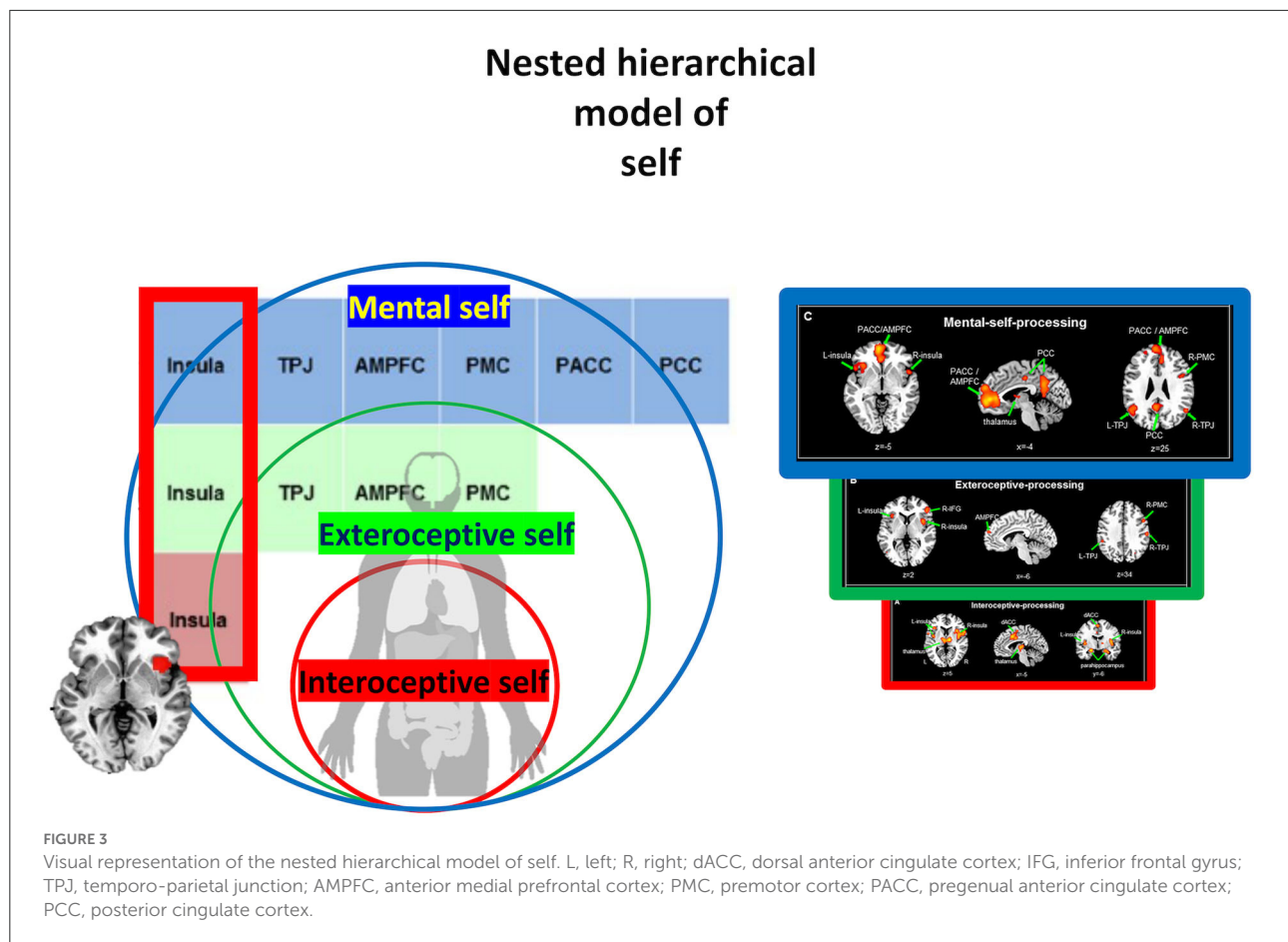


a hierarchy of the self in the brain with bodily self, autobiographical self, and extended self.

Damasio (2010) in his model theorized the “proto self,” which generates primordial feelings and is largely subconscious. This bottom layer of self provides the grounding for the higher levels, such as the “core self,” which represents the transient relationship between an individual and the surrounding environment, and finally, the “autobiographical self,” which requires working and long-term memory.

A more radial-concentric approach to the brain is the three-layer anatomical model of the brain as proposed by Feinberg and Northoff (Northoff et al., 2011). These proposals parallel recent socio-psychological and phenomenological findings (Gallagher, 2000; Nelson et al., 2014) and support the evidence for a functionally constituted entity of the self ranging over multiple interacting levels from an unconscious, pre-reflective, and minimal self to a reflective phenotypic “idiographic” narrative

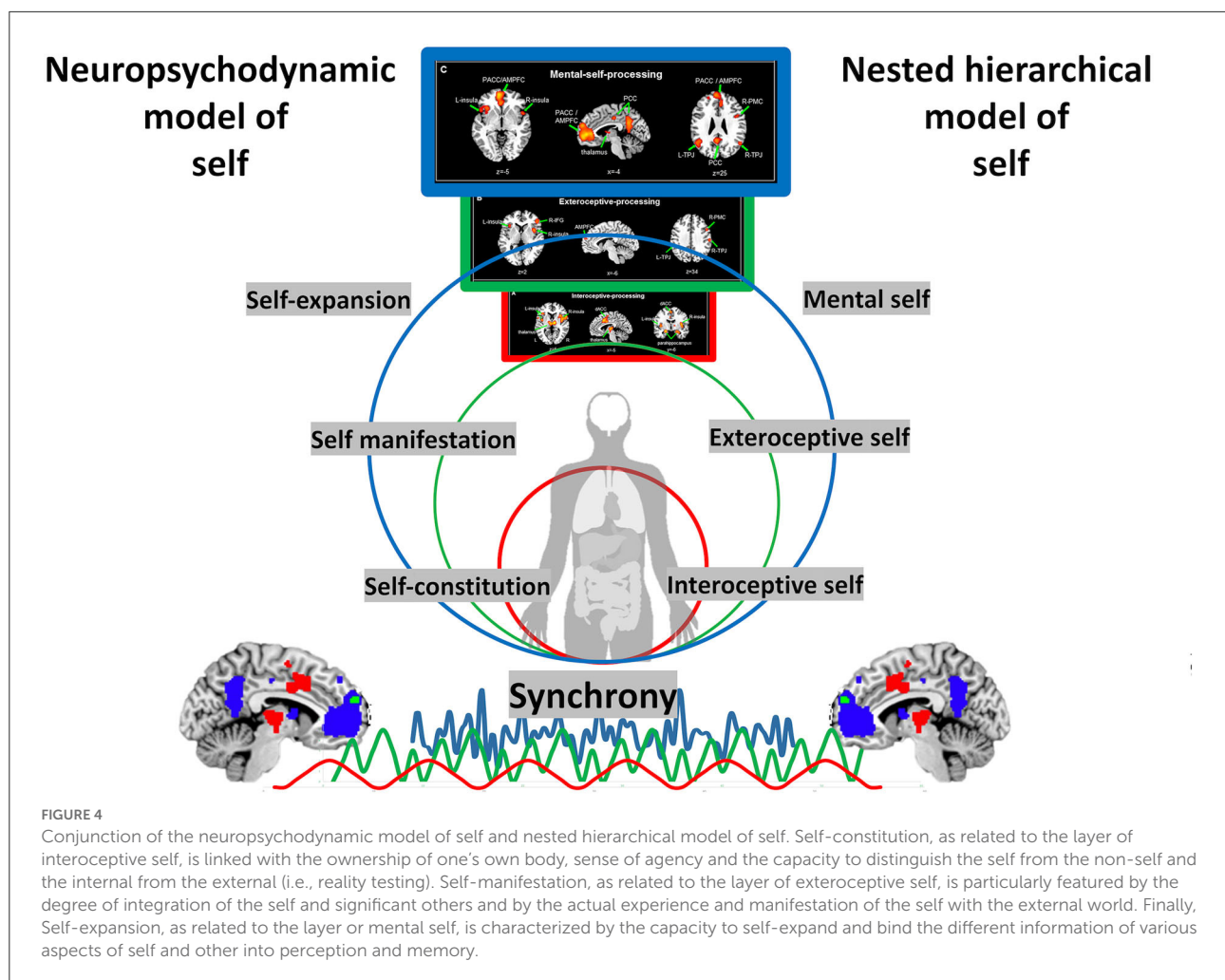
self constituted by interpersonal and sociocultural experiences. These proposed hierarchies of self seem to be in line with Endel Tulving’s three-layer theory of consciousness (Tulving, 2002): *anoetic* (unthinking forms of experience, which may be effectively intense without being “known”) *noetic* (thinking forms of consciousness, linked to exteroceptive perception and cognition) and *autonoetic* (abstracted forms of cognitions and perceptions and cognitions, which allow conscious “awareness” and reflection through episodic memories and fantasies. These layers, accordingly with Solms and Panksepp (2012), seem to be connected with some major evolutionary passages of the brain, such as (i) the evolution of the upper brainstem (up to the septal area) as related to *anoetic consciousness*, (ii) the evolution of lower subcortical ganglia and upper limbic cortical midline structures as related to *noetic consciousness*, and (iii) the evolution of higher neocortical functions and association cortices as related to *autonoetic consciousness*.



A recent evidence-based, large-scale fMRI meta-analysis in healthy subjects suggests a multilayered nested hierarchical model of self (Qin et al., 2020) including: (i) interoceptive self; (ii) extero-proprioceptive self; and (iii) mental self.

The interoceptive self refers to the processing of the body's inner organs and was investigated through an fMRI task related to interoceptive awareness of one's own body. The extero-proprioceptive self focuses on external or proprioceptive bodily inputs and was investigated with fMRI studies focusing on external bodily-related inputs, such as facial or other proprioceptive inputs. Finally, the mental self was investigated considering all task employing trait adjectives or other self-related stimuli vs. non-self. Intriguingly, the studies related to the interoceptive self emphasizes the role of bilateral insula, dorsal anterior cingulate cortex, thalamus, and parahippocampus, which are also considered core regions of the salience network (Menon and Uddin, 2010). The extero-proprioceptive self yielded regions, such as the bilateral insula, inferior frontal gyrus, premotor cortex, temporo-parietal junction (TPJ), and medial prefrontal cortex (MPFC). These regions share the processing of proprioceptive inputs related to the body, and this seems to be closely related to the

concept of “embodied self” (Gallagher, 2005; Tsakiris, 2017). Finally, the mental self, related to fMRI studies, yielded DMN cortical midline regions, such as the medial prefrontal cortex and posterior cingulate cortex, as well as the regions included in the extero-proprioceptive self, most notably the bilateral TPJ, as well as regions of the interoceptive self, i.e., bilateral insula and thalamus. Together, these findings describe a hierarchical model of self (Qin et al., 2020) showing how regions of the interoceptive self are also included in the other layers (extero-proprioceptive and mental self) where they were complemented by additional regions extending the topography of the self. Moreover, this study also highlights the function of the insula, which seems to represent the common denominator for each level of self processing (refer to Figure 3). This hypothesis linking the functional role of the insula and the different layers of the sense of self seems to be further confirmed by the studies on macaques (Critchley and Seth, 2012), studies on the role of the insula and emotional awareness (Gu et al., 2013) and by its relevance in severe psychopathologies, such as anorexia (Esposito et al., 2018) and anxiety disorders (Lucherini Angeletti et al., 2021).



This nested topographical hierarchical model of self might be associated with and add more information to the proposed neuropsychodynamic nested model of self (Scalabrini et al., 2018).

In this context, the authors propose a multilayered model of the self-departing from the building blocks of relational alignment to the different layers of self, named: (i) self-constitution; (ii) self-manifestation; and (iii) self-expansion, which might somehow parallel the: (i) interoceptive; (ii) extero-proprioceptive; and (iii) mental self (refer to Figure 4)

As proposed by the authors (Scalabrini et al., 2018), relational alignment or synchronization is considered the prerequisite that gives the newborn the framework for the sense of subjectivity that is dependent on the first encounter with the other (e.g., the caregiver). The encounter with the other may facilitate (or not) the constitution and development of the self, depending on the degree of attunement that plays a fundamental role in shaping the sense of self, relatedness, and the capacity to regulate emotions and to mentalize among the complexity of

psychological development (Schore, 2001a,b, 2012, 2021; Mucci, 2013, 2018a, 2021a,b).

Intriguingly, aging and dysfunctions have been related to decreased neuronal complexity in regions typically involved in “more cognitive” task (e.g., frontal and parietal lobes), while increased neuronal complexity has been found in regions, such as insula, limbic, and temporal lobe, typically considered as “emotional” and/or more specific to the self (Cieri et al., 2021). These findings further support the relevance of the self in development and aging of the individual.

As previously shown, this hypothesis is further supported by literature showing how the child's development and their growing brain are optimized where the provision of parental care is sensitively attuned to the infant's needs (Atzil and Barrett, 2017; Atzil et al., 2018 for a perspective). Atzil et al. (2018) propose how the insula and the regulation of interoception and allostatic needs of the child might play a fundamental role in the development of the self and of a social brain.

The emphasis on the interoceptive and allostatic needs of the child further supports the hierarchical model of self (Qin et al.,

2020) and the importance of the insular cortex as a crossroads for the relation integration between internal and external stimuli (Craig, 2010; Menon and Uddin, 2010).

It seems that these synchronous relational aspects, together with the key role of the insula, are the prerequisite for the three layers of the self. Intriguingly, in a recent article investigating the spatiotemporal dynamics of the three layers of self in the spontaneous activity of the brain (Scalabrini et al., 2021b), the right anterior insula exhibited increasing centrality indices over the three hierarchical layers of self in comparison with all other regions implicated. High centrality and functional connectivity indicate that the right insula synchronizes and integrates the activity of the other regions of the self-networks. Functional connectivity allows for functional integration, that is, the degree to which a region pools or sums the activity of other regions within its own neural activity through synchronization (Deco et al., 2015). The increased high degree of centrality over networks enables the right insula to integrate interoceptive, exteroceptive, proprioceptive, and cognitive/mental functions linking the three layers of self-specific information which are thereby spatially or topographically nested within each other (Qin et al., 2020).

Moreover, the right insula exhibits longer time windows in its neural activity than both the left insula and the other regions of the different layers of self. Together, these findings suggest higher degrees of both functional and temporal integration in the neural activity of the right insula. This suits the right insula to ideally serve as a topographic and dynamic node or glue between the distinct layers of self and their high degrees of spatial nestedness and temporal continuity.

Taken together, here, we provide a parallel between these two models that give us a neuropsychodynamic topographical grounded model of self which carries major psychodynamic and neuroscientific implications.

We hypothesize that the right insula's key role in constituting spatial nestedness on the neuronal level, i.e., among the three layers of self-networks may also be in manifesting on the psychological level: self-specific interoceptive information may be contained and nested within the layer of self-specific proprioceptive and exteroceptive information which, in turn, may be nested and contained within the even more extended layer of self-specific mental or cognitive information (refer to Figure 5).

From self to trauma – topographic-dynamic layers of trauma

We now assume that the nested hierarchy of self might be connected to the hierarchy of trauma, that is, the effects traumatic events can induce in the self and ultimately

the brain. Our observations follow Clara Mucci's excellent account of trauma through interdisciplinary and clinical observation (Mucci, 2013, 2014, 2018a, 2019, 2021a,b, 2022): she distinguishes three layers of inter-personal trauma which we, in a second step, connect with the different layers of self in its nested neural hierarchy.

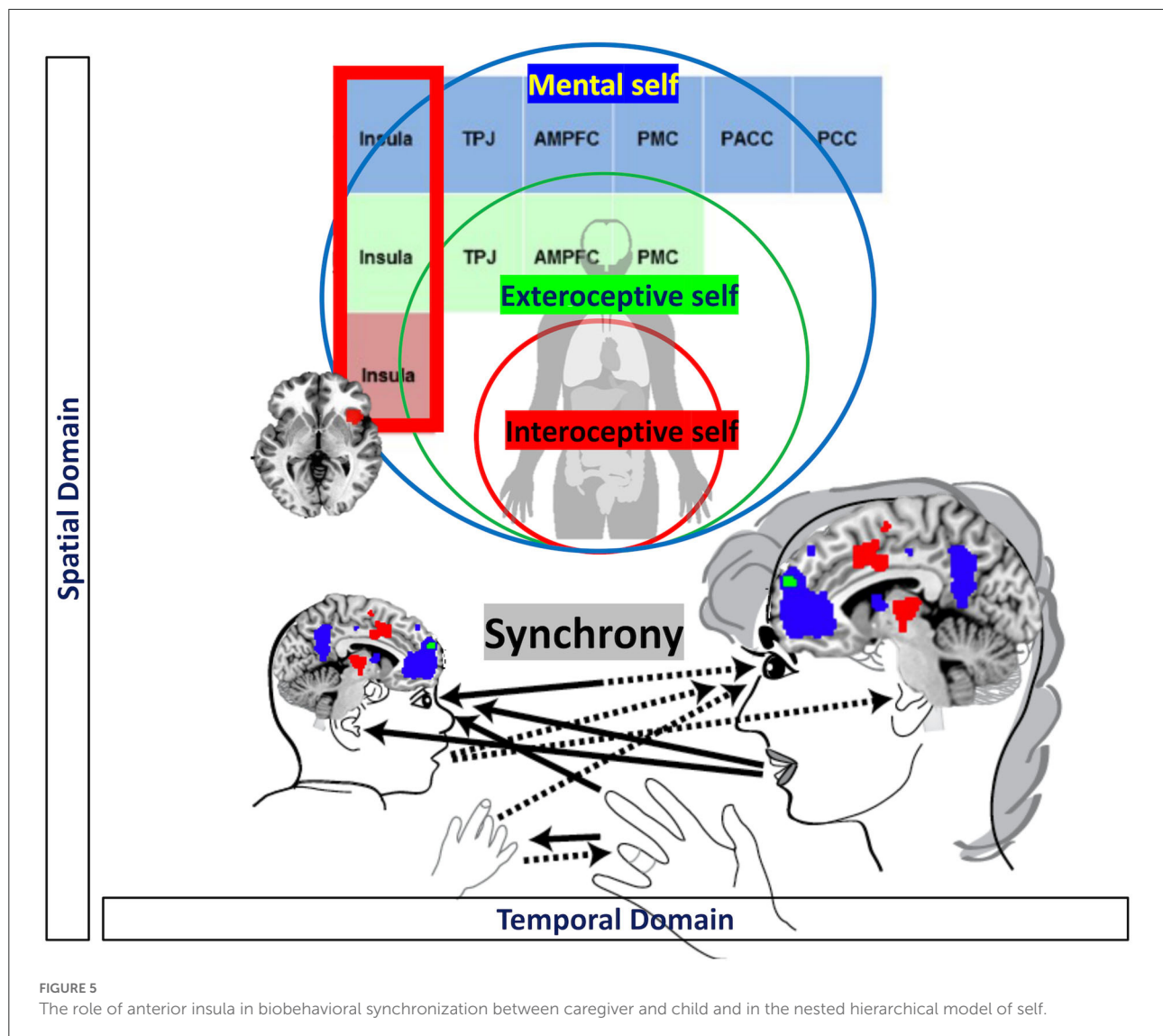
Mucci (2013, 2018a, 2022) distinguishes between three levels of inter-personal trauma as distinct forms and different degrees in the severity of disrupted attachment across the life span from infancy to adulthood. Notably, all three levels are not separated entities but rather refer to different degrees of an underlying continuum of trauma severity and depth. Hence, the distinction of the three levels is more heuristic, serving conceptualization; they should not be confused with the actual reality itself where one can observe multiple transitions and overlaps with possible cumulative effects and influences between one level and the other.

First layer of trauma I – early relational trauma, subcortical regions, and interoceptive self

The first, most basic level of inter-personal trauma concerns what Mucci (2013, 2018a), and Schore (2003a,b, 2009) describe as “early relational trauma.” This primarily goes back to early infancy, where the infant suffers from a lack of attunement and synchronization with the caregiver, who is unable to provide secure, stable, and continuous care and containment for the infant (Schore, 2003a,b). This creates disorganized attachment (Liotti, 1992) in the infant and makes it prone to most basic dissociative responses (Schimmenti and Caretti, 2016; Farina et al., 2019; Scalabrini et al., 2020a; Mucci, 2021a,b; Mucci and Scalabrini, 2021) in childhood and later in adulthood. Moreover, the infant or child suffers from affective dysregulation with hypoarousal, which later may translate into high vulnerability for depression, anxiety, and the development of trauma-related personality disorders.

Neuronally, the early relational trauma has been associated with various cortical and subcortical brain regions that belong to the limbic system, including the insula, the amygdala, and the orbitofrontal cortex (Siegel, 1999; Schore, 2003a,b). Interestingly, these regions strongly overlap with those recruited during the most basic layer of self in the nested neural hierarchy, the interoceptive self: subcortical regions such as the thalamus and those in midbrain and even brain stem, as well as cortical regions such as the insula mediate this most basic layer of the interoceptive self.

These regions predominantly process and integrate interoceptive stimuli and inputs from the subject's own inner body and, in later stages, their integration with exteroceptive inputs. This is the most fundamental layer of the self which



ties it closely to the body and environment, i.e., the most basic intero- and exteroceptive inputs, any infant (and later child and adult) receives. Disturbances in these regions' interoceptive processing consequently lead to major disturbances in the most basic sense of self, i.e., the interoceptive self.

First layer of trauma II – from irregular/absent interoceptive input to disordered attachment

How is this most basic and fundamental layer of the interoceptive self related to the symptoms of early relational trauma? If the caregiver like the mother does not provide proper basic and affective care, such as nutrition and emotional

mirroring, in a regular and predictable way, the interoceptive inputs to the subcortical regions and the insula are highly irregular and/or absent. What is described as “*disorganized attachment*” on the psychological level (Liotti, 1992; Schore, 2003a,b, 2009; Mucci, 2013, 2018a, 2022) may then be related to “disorganized interoceptive input processing” on the level of the brain (refer to Atzil et al., 2018): irregular or even absent interoceptive inputs will make it rather difficult for these limbic subcortical and cortical regions to develop a proper dynamic, that is, in neuronal spatiotemporal terms a power spectrum with the balance of slow and fast frequencies in a scale-free and temporally well integrated and nested way.

Specifically, there may be a disbalance of slow and fast frequencies which, in turn, impedes all subsequent processing in these limbic subcortical nuclei and related cortical regions, including their integration of the various interoceptive

(and later also exteroceptive) inputs. If, however, intero-intero/exteroceptive integration is impaired, the sense of their interoceptive self may become unstable and fragile – the self of those subjects shows a lack of energy with hypoarousal (as eventually related to the Nucleus Basalis Meynert), affective dysregulation (as eventually related to serotonergic and adrenergic/noradrenergic subcortical regions; refer to [Panksepp and Biven, 2012](#)), and dissociation proneness (as related to the lack of intero- and exteroceptive integration and decreased connectivity; [Scalabrini et al., 2020a](#)).

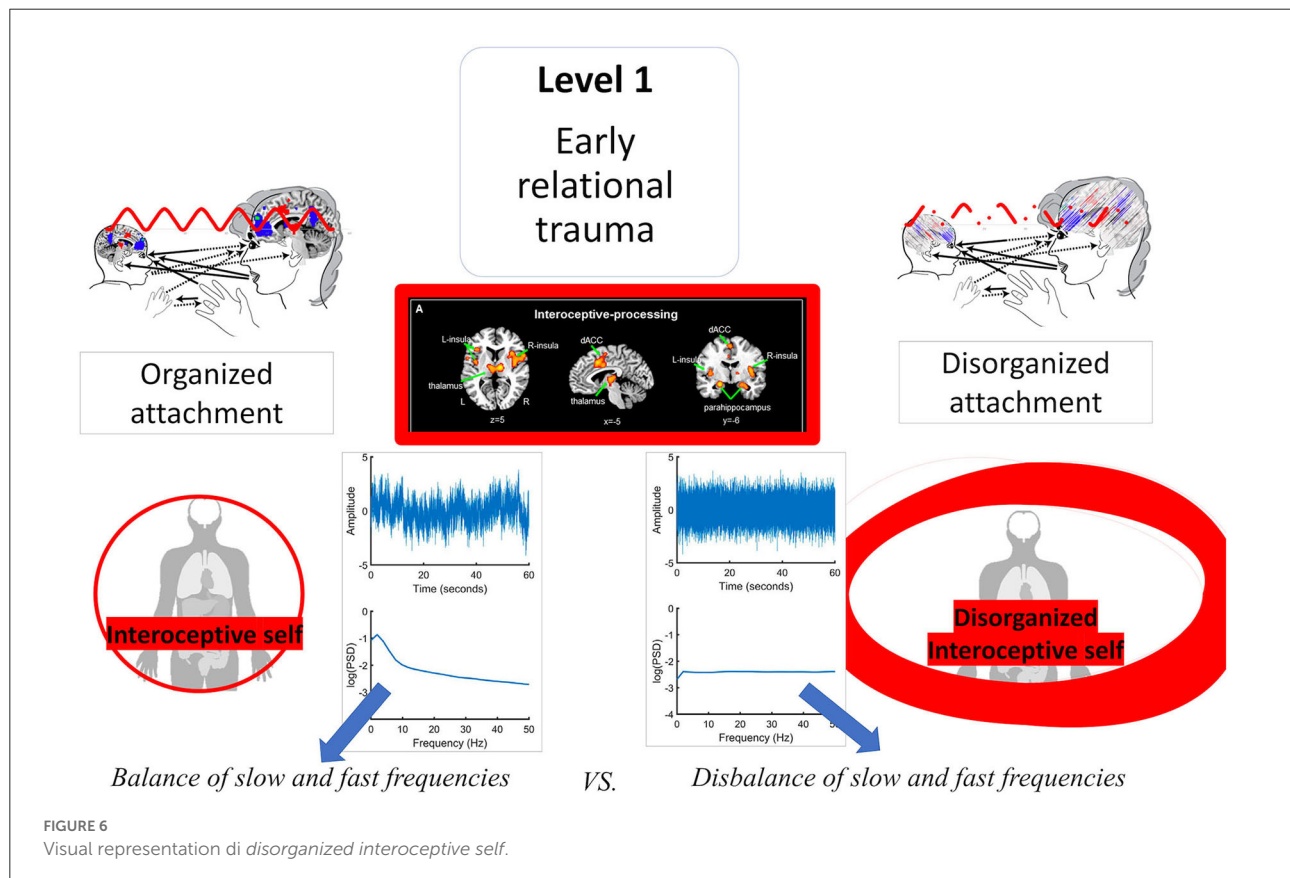
Together, irregular or absent interoceptive inputs through the caregiver may translate into corresponding neuronal topographic disbalances and dynamic irregularities in the power spectrum predominantly in limbic subcortical-cortical regions. The subcortical-cortical regions' abnormally slow-fast balanced power spectrum with decreased temporal nestedness, in turn, may be related to decreased arousal, affective dysregulation, and an unstable or fragile interoceptive self. These processes manifest in what has been described as “disorganized attachment” on the psychological level ([Liotti, 1992](#); [Mucci, 2018a](#)) that neurally may be reflected in the brain's “disorganized topography and dynamic” as the basis for the so-called psychological baseline and social brain ([Northoff, 2016](#); [Atzil et al., 2018](#); [Scalabrini et al., 2022](#); refer to [Figure 6](#)).

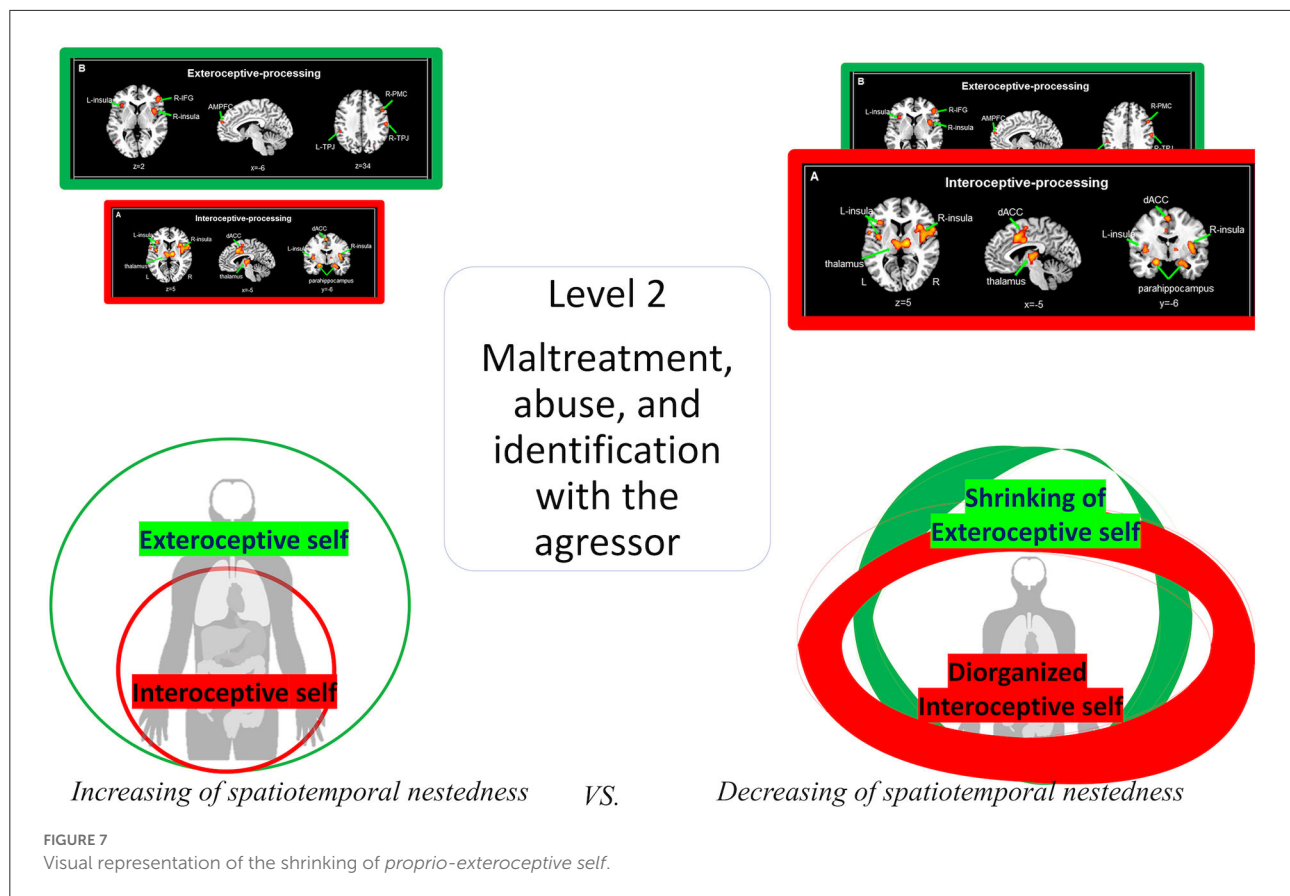
Second layer of trauma I – from maltreatment and abuse to the fragmentation of self

The second level of inter-personal trauma concerns what [Mucci \(2018a\)](#) describes as “maltreatment and abuse.” This may concern events that often exceed beyond the “normal” ranges of human experience like maltreatment, severe emotional and physical deprivation, incest, and abuse. These events can occur at every time of the life cycle, including in infancy or early childhood where they may exert major reverberation for the psyche.

As focused here, these events may mainly occur in infancy and early childhood where they leave major consequences for later adulthood. Dissociation and/or identification with the aggressor characterizes this second level of trauma described by [Mucci \(2013, 2018a, 2022\)](#) which adds on to the first level, the non-intentional disattunement of the caregiver and places additional stress and dynamic consequences of severe abuse or maltreatment and severe deprivation (and the identification with the aggressor, with the internalization of guilt and rage, often revolved around the self).

In the research, maltreatment and abuse have been indicated as a vulnerability to develop disorganized attachment





in childhood and possible dissociative behaviors (Main and Hesse, 1990; Liotti, 1992; Carlson and Sroufe, 1995; Draijer and Langeland, 1999; Ainsworth and Eichberg, 2006). In addition, it creates a dissociated personality and internalized victim/persecutor dyad (refer to Kernberg, 1967; Fonagy and Bateman, 2007; Mucci, 2018a) where one identifies with the victim and the related effects (low self-esteem, blame, shame, guilt), yet also identifies with as an internal persecutor position of which the subject is not aware, and identifies with the effects of the persecutor which include violence, hate, and aggressiveness (often revolving around one's body). The identification with the aggressor (Ferenczi, 1932a/1988,b) is a way that the personality adopts to survive. The internalized persecutory parts can act against oneself (damaging oneself and one's resources) or against the other (becoming actually violent against another individual).

The second layer of the nested hierarchy of self therefore extends beyond the interoceptive self to its proprio-exteroceptive input from the outside of the own body and others' relative to the own body.

This is neurally reflected in the recruitment of regions, such as the anterior medial prefrontal cortex, premotor, fusiform face area, and temporo-parietal junction, that typically process these kinds of inputs

(refer to above and Qin et al., 2020). Importantly, this proprio-exteroceptive layer of self builds and nests upon the more fundamental first layer, the interoceptive input layer of self: the latter and its regions are integrated within the former. Such nested hierarchical relationship carries major implications for how the self can cope with trauma (refer to Figure 7).

Second layer of trauma II – topographic-dynamic re-organization of brain and self

One may now postulate that the second level of trauma with abuse, maltreatment, and/or physical and emotional deprivation may primarily affect the extero-proprioceptive layer of the self by disrupting the activity of its respective regions. If that extero-proprioceptive input layer is disrupted, neural activity will “fall back” onto the more basic and fundamental layer, the interoceptive input layer.

At the same time, the extero-proprioceptive layer of self may be split off from the nested hierarchy, including its containment or nesting within the interoceptive layer of self. In a way similar to Russian dolls, such that this is like

“taking out” one of the more intermediate-sized Russian dolls which then, relatively speaking, will enlarge the next larger one in a disproportionate way. Though, one needs to be more careful in this interpretation. The traumatized layer is not completely “taken out” as it is still present as described by Ferenczi and still manifests in the various affective-emotional symptoms related to the trauma. In this case, the trauma leaves its traces at a bodily interoceptive level. In contrast, the extero-proprioceptive level, instead of being literally “taken out,” “shrinks” in an abnormal way whereby it may more or less disappear into the previous one, the interoceptive layer.”

Psychologically, this means that the extero-proprioceptive self may become abnormally small, shrink, and no longer be visible within the more overarching interoceptive layer of self. Put in spatiotemporal terms, this means that the spatial extension and the temporal range of the extero-proprioceptive self may be reduced to the ultimately smaller spatial and temporal ranges of the interoceptive self. The nested hierarchy of the self and brain thus may be disrupted in its degree of temporal and spatial nestedness between the different layers. The extero-proprioceptive self may no longer be perceived as related to the self, thus creating the circumstances for aggressive behaviors manifested toward the bodily self or toward the external world.

Trauma may then be understood as disruption of the temporo-spatially nested hierarchy of both the brain and self.

How can the person defend its own self and react to its decreased temporo-spatial nestedness of its different layers of self? How is such “filling of the extero-proprioceptive missing” manifested on the psychological level?

As theorized by Mucci (2013, 2018a), we can see the victim/persecutor split in severe personality disorders, which are especially common in young adults nowadays.

On the one hand, we can observe how borderline patients self-cut or try to commit suicide attacking their own body subjugated by the disorganized interoceptive self. Contrastingly, severe narcissists rely on the mental effort to maintain a sense of pathological grandiose self which damages their own resources (e.g., time and intelligence) to protect their own inner fragility and low self-esteem (for an fMRI study on narcissistic grandiosity refer to Scalabrini et al., 2017). The identification with the aggressor, with the internalization of the rage against one's body or against the other, explains much of the destructiveness against self and others in personality disorders and other severe mental pathology (Mucci, 2013, 2017, 2018a).

In sum, the changes related to the second layer of trauma reflect the re-organization of (i) the spatial topography of the nested neural hierarchy of self and (ii) its related power spectrum with its slow-fast balance. Hence, what we, from the outside, describe as disruption of the temporo-spatial nestedness of the neural hierarchy of self may, from the subject's inside or its first-person perspective, be a topographic-dynamic

re-organization of its own self to cope with the trauma albeit being ultimately maladaptive.

Third layer of trauma – from massive trauma to topographic-dynamic re-organization of the mental or cognitive self

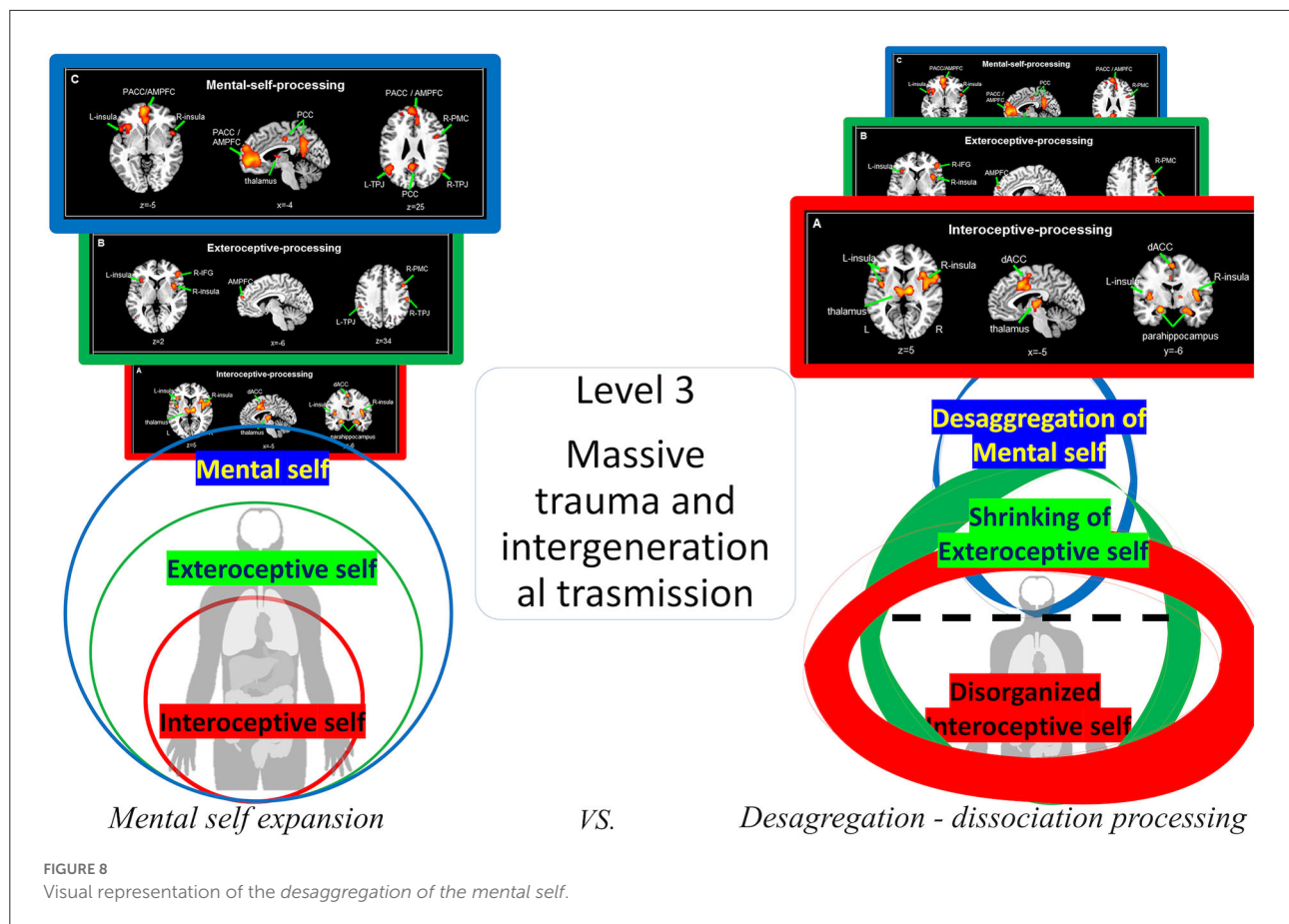
There are events, such as genocide, war, and rape/abuse, that lie outside the “normal” range of human experiences. How does the brain and self react to such cumulative massive traumatic experiences?

This depends first on the intrinsic factor like the stability of the nested hierarchy of the self and second on the severity of traumatic occurrence and its cumulative effects. The same trauma may have different impacts depending on whether it occurs during infancy, adolescence, or adulthood as these are related to the different layers of the nested hierarchy of self. It also depends upon the severity of trauma and the history of cumulative trauma experiences by the subject: the more severe a trauma and the more severe the history of traumatic experiences in the past (1st and 2nd levels), the deeper it will reverberate into the deeper layers of self disrupting its own stability.

One may assume that massive traumata primarily affect the most upper layer of self, the mental or cognitive self, and its underlying regions such as the cortical midline structures. Neural activity in these regions may be diminished and reduced in the face of such severe trauma which is indeed supported by various results (Mucci, 2018a, 2021a,b; Scalabrini et al., 2020a; Mucci and Scalabrini, 2021). The self as mental baseline or default mode of the brain (Northoff, 2016; Scalabrini et al., 2018, 2021a) is put “out of order.” The mental self as the most upper layer of self is consequently diminished.

This, in turn, changes the whole nested hierarchy of self. The most upper layer, the mental self, is now abnormally “shrunk” which, relatively speaking, renders the interoceptive and extero-proprioceptive layers of self abnormally strong. This is, for instance, reflected in the often observed increase of anxiety with extreme interoceptive awareness in these subjects that often suffer from various psychopathological syndromes, such as depression, anxiety, trauma-related, and personality disorders (Mucci, 2018a; Scalabrini et al., 2018, 2021a).

At the same time, the mental or cognitive features of the self may be “split” from its lower nesting layers, the intero- and extero-proprioceptive layers of self. That may be reflected in the various compartmentalization and detachment symptoms of dissociation (Scalabrini et al., 2020a). Dissociation thus may be understood as the disruption of the “glue” or nestedness among the different layers of self – this, as we postulate, may be traced to a corresponding disruption in the nestedness of the neural layers of self (see Figure 8).



Dissociation as a disruption of temporo-spatial nestedness

Dissociation plays a key role in trauma which, as we assume, connect at the neural and mental levels. What is described as dissociation of parts of the self on the psychological level may be mediated by a corresponding dissociation on the neural level of self: the neural level of the default-mode network mediating the mental or cognitive layer of self dissociates from its nesting neural layers, the subcortical-cortical layers of both interoceptive and extero-proprioceptive self. Correspondingly, the dynamics of the power spectrum will dissociate in some of its parts, such as its slower and/or faster frequency ranges. This results in the disruption of the temporo-spatial nestedness of the brain's hierarchical topography and dynamics. As in the case of the second layer of trauma, subjects may react to massive trauma by topographic-dynamic re-organization of their brain and its self by strengthening the respectively preserved layers.

What happens when the subjective synchronous experience with the self, the body and the world is disrupted?

In a recent article by Scalabrini et al. (2020a), the central role of the right anterior insula is considered as the central hub for the temporo-spatial synchronization between body

and brain, that is disturbed in dissociation. Such temporo-spatial desynchronization of the body-brain connection should disrupt the first-person perspective (Tallon-Baudry et al., 2018): instead of being perceived in relation to the own person (in a first-perspectival mode), the contents would remain detached from the self and its bodily-based first-person perspective: the contents will then be perceived as such but in such a way that they remain unrelated to the person and its bodily-based first-person perspective – the contents will consequently be perceived as foreign or alien, e.g., as non-self-related, thus reflecting what, on the symptomatic level, is described as detachment symptoms.

More specifically, the desaggregation of intero-exteroceptive function based on the impaired spatiotemporal integration leads to the alterations in embodiment and, even further, the disruption of our self's relationship with the others and the environmental context. This failure of integration can thus be seen as a disconnection that induces an instant collapse of both subjectivity and intersubjectivity, supporting affective dysregulation (Schore, 2001a,b, 2003a,b; Mucci, 2018a; Schimmenti and Sar, 2019; Scalabrini et al., 2020a,b; Cavicchioli et al., 2021). Such disruptions of embodiment and first-person perspective are well reflected in trauma-related disorders. Stressful stimuli, especially

those associated with painful emotional effects, are thus not experienced in consciousness, and they are associated with what Bromberg (2014) terms “not-me” self-states. At the same time, intero-exteroceptive desegregation leads to disruption of the self’s attunement with the other and consequently the world. This reinforces a vicious circle of traumatic experience that, through dissociation, are intergenerationally transmitted between caregivers and child (Mucci, 2013, 2018a, 2022) (see Figure 9).

Conclusion: Synchrony between and beyond layers – topographical and dynamic re-organization of the nested hierarchy of self and its biobehavioral-affective regulation

We characterize the effects of trauma by an abnormal shift in the topography and dynamic of the nested hierarchy of self. The trauma-related symptoms are conceived here as topographic-dynamic re-organization of self and brain albeit in

a maladaptive way, that we call the traumatic re-organization of the nested hierarchy of the self (see Figure 10 for a summary).

As we recently proposed (Northoff and Scalabrini, 2021), the aim of a neuropsychodynamic informed psychotherapeutic process is to (i) to reverse such maladaptive topographic-dynamic reorganization of brain and (ii) to establish a more adaptive and stable temporo-spatial nestedness of brain and self thereby re-establishing a proper nested hierarchy of self. This process might serve to re-establish the subjective sense of integrity, coherence, and continuity of self over time and space, “the capacity to feel like one self while being many” (Bromberg, 1996, p. 1. 166)

Moreover, to re-establish and integrate the sense of self of the patients, the therapist should act as a regulator of different affective and self states working at the edges of affective dysregulation and in-between different states of mind (Mucci, 2013, 2018a, 2021a,b; Mucci and Scalabrini, 2021). As Schore writes: “with early forming severe right-brain pathologies, the clinician’s primary function is as an affect regulator for the patient’s primitive, traumatic states, including those affective states that are walled off by dissociation” (Schore, 2003a,b p. 246).

Desynchronization – Desaggregation of neuronal and psychological processing

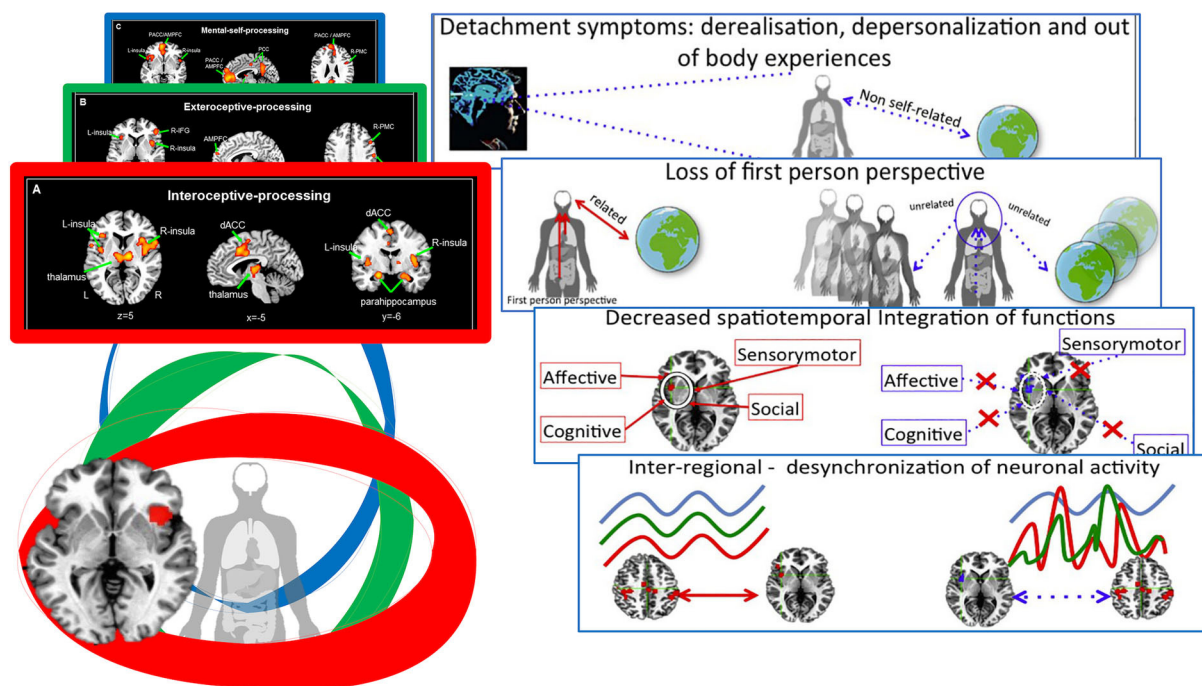
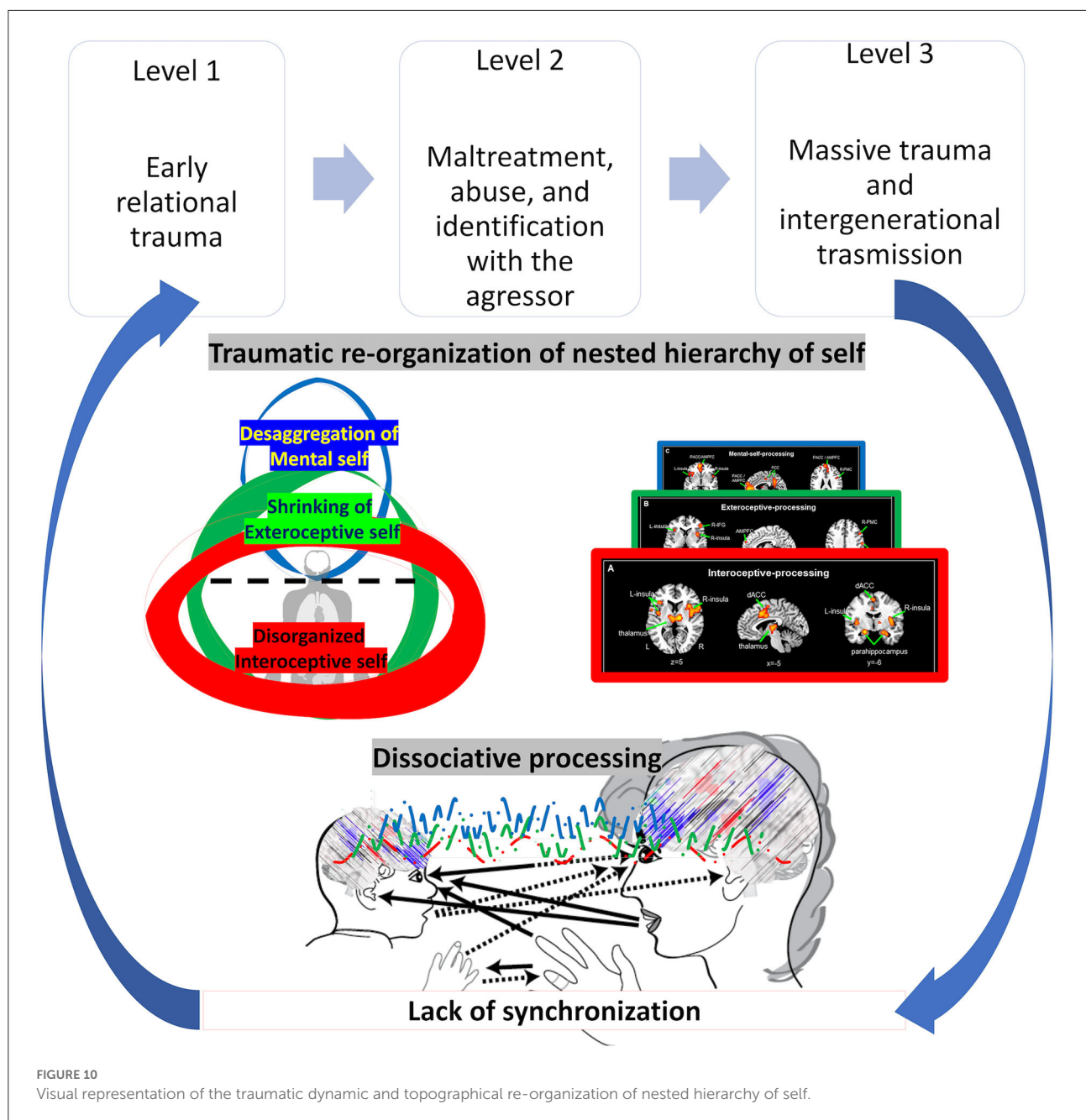


FIGURE 9

Visual representation of dissociation and traumatic re-organization of nested hierarchy of self.



Dissociation, as we suggested before, operates in terms of lack of integration and synchronization between the different layers of the self (Scalabrini et al., 2020a). As a consequence, moving beyond dissociation means to re-establish the nested hierarchy of the self system to process at the mental level (mental self) the perception of external stimuli (proprio-exteroceptive self, i.e., information coming from the relational environment) and integrate them with internal stimuli (interoceptive self, i.e. information from the body, the “felt experience” - Craig, 2010, 2011).

Our claim seems to perfectly fit to what Mucci (2021a,b) and Schore (2011) describe in terms of dissociative processing:

dissociation thus reflects the inability of the right brain cortical sub-cortical implicit self system to recognize and process the perception of external stimuli (exteroceptive information coming from the relational environment) and on a moment-to-moment basis integrate them with internal stimuli (interoceptive information from the body, somatic markers, the ‘felt experience’). This failure of integration of the higher right hemisphere with the lower right brain

system and disconnection of the central nervous system from the autonomic system induces an instant collapse of both subjectivity and intersubjectivity. Stressful effects, especially those associated with emotional pain, are thus not experienced in consciousness, (Bromberg 'not-me' self-states). (Schore, 2011, p. xxxiii)

Consequently, healing the self and re-establishing the sense of self-continuity beyond the dissociation of its trauma may primarily require one to re-establish or re-organize the topography and dynamic of the nested hierarchy of self and its brain through spatial and synchronic means. Synchronization here refers to what Feldman terms as “social synchrony” in mother-child dyads: “a move from a focus on one-brain functioning to understanding how two brains dynamically coordinate during real-life social interactions” (Levy et al., 2017, p. 1036). This is similar to what we defined as “relational alignment” (Scalabrini et al., 2018) and Schore (2021) described as “interbrain synchrony” referring to an alignment of brains between two individuals, that is the “simultaneous changes of emotional energy within the brains of both members of the dyad. This mutual entrainment of brain during moments of synchronization triggers an amplified energy flow, which allows for a coherence or organization of self and its implicit and explicit conscious experience.” (Schore, 2021).

At a psychotherapeutic level, this seems to resemble to what has been theorized as “critical or now moments in psychotherapy” by Stern (2004) and the Boston Change Process Study Group (2005, 2013, 2018). For the BCPSG, the goal in psychotherapy is to share similar mental landscapes to understand and be understood for the re-organization of the sense of self and relatedness. This intersubjective sharing includes both the explicit verbal meaning of what one says and the implicit nonverbal meaning and in any event in time, “intersubjective sharing” is the primary goal of the therapy. The units of interaction are called relational moves and the goal of relational moves is to adjust or regulate the “intersubjective field,” that is, the shared mental/feeling landscape. These moves can consist of a spoken phrase, a silence, a gesture, or shift in posture, or a facial expression. Moving along can lead to sudden dramatic therapeutic changes by way of “now moments” and “moments of meeting.” That in our terms might be considered as moments of desynchronization and moments of synchronization. These moments shape the intersubjective field that gets suddenly reorganized: this occurs when the current state of implicit relational knowledge is sharply thrown into question and basic implicit assumptions about the relationship are now put in the discussion.

There is a mounting affective charge, which can be more or less regulated. The situation emerges unexpectedly and something must be done (including the option of

doing nothing). Ideally following these critical moments “moments of meeting” or new moment of synchrony might take place to resolve the crisis. The moment of meeting seeks to use the disorganization of the now moment to enlarge the intersubjective field in ways not thought of before to reconnect the different layers of self. The therapist needs to synchronize (virtually or symbolically) her/his larger (spatial-topographic and temporal-dynamic) scales of her/his nested hierarchy of self to the more restricted of his client’s traumatized self. These moments of meeting or synchrony require an authentic response finely matched to the momentary local situation to provide the client with the opportunity to integrate and nest her/his own more restricted spatiotemporal scales of her/his interoceptive self in a virtual, i.e., inter-personal way into the larger ones of her/his therapist.

A moment of meeting or synchrony requires an authentic response finely matched to the momentary local situation. It must be spontaneous and carry the therapist’s personal signature. In that way, it reaches beyond a technical, neutral response and becomes a specific fit to a specific situation. These moments have been lived through together. That, in turn, will create the basis of a “psycho-physiological secure baseline” (refer to Northoff, 2016; Northoff et al., 2022; Scalabrini et al., 2022) to maintain a continuous and temporally extended and integrated sense of self and sense of relatedness that will allow the client to process the traumatic input relationships in a non-threatening and non-disrupting way for her/his own self way without becoming fragmented or dissociated. This result in a re-organization of the border between order and chaos, of coherence and complexity. The therapist and the patient have created an expanded intersubjective field that opens up new possibilities of ways of being with one another where the traumatic input relationships associated with the own interoceptive self are now integrated and nested virtually (or symbolically) within the therapists’ larger spatiotemporal scales.

Practically speaking, the use of synchronic now moments in psychotherapy involves spontaneity and authenticity. Here, we propose to use temporo-spatial coordinates in the psychotherapeutic setting to work within the transference-countertransference matrix of the therapist–patient dyad. Working through synchrony, space, self-relatedness, and affective regulation might provide a more comprehensive, basic, and extensive operating field that also embeds and contains affective, social, cognitive functions within a larger more comprehensive context. This process aims to shape the temporal dynamic flow of the patient’s neural and psychic activities to re-organize the traumatized self and consequently change the patient’s baseline organization of the self and its spontaneous psychic and neuronal activities (Northoff and Panksepp, 2008; Scalabrini et al., 2018, 2022; Mucci, 2021a,b; Northoff et al., 2022).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This project/research was supported by “Search for Excellence – UdA” (University G. d’Annunzio of Chieti-Pescara) to AS for the project SYNC (The Self and its psychological and Neuronal Correlates – Implications for the understanding and treatment of depression as a disorder of Self) and by the Michael Smith Foundation Canada Research Chair, by the grant from the Ministry of Science and Technology of China, National Key R&D Program of China (2016YFC1306700) and from the European Union’s Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement

No. 785907 (Human Brain Project SGA2), the ERANET grant, NFRF grant, and Team grant from uOMBRI to GN.

Acknowledgments

The authors would like to thank Yasmin Harrington and Halsey Keiler for reviewing and editing the manuscript for language and grammatical issues.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abraham, E., Hendler, T., Zagoory-Sharon, O., and Feldman, R. (2016). Network integrity of the parental brain in infancy supports the development of children’s social competencies. *Soc. Cogn. Affect. Neurosci.* 11, 1707–1718. doi: 10.1093/scan/nsw090
- Ainsworth, M. D. S., and Eichberg, C. (2006). “Effects on infant-mother attachment of mother’s unresolved loss of an attachment figure, or other traumatic experience,” in *Attachment Across the Life Cycle* (Routledge), 168–192.
- Alcaro, A., Carta, S., and Panksepp, J. (2017). The affective core of the self: a neuro-archetypal perspective on the foundations of human (and animal) subjectivity. *Front. Psychol.* 8, 1424. doi: 10.3389/fpsyg.2017.01424
- Apps, M. A., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41, 85–97. doi: 10.1016/j.neubiorev.2013.01.029
- Arrieta, M. C., Stiensma, L. T., Amenyogbe, N., Brown, E. M., and Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Front. Immunol.* 5, 427. doi: 10.3389/fimmu.2014.00427
- Atmanspacher, H. (2012). Dual-aspect monism à la Pauli and Jung. *J. Conscious. Stud.* 19, 96–120.
- Atzil, S., and Barrett, L. F. (2017). Social regulation of allostasis: Commentary on Mentalizing homeostasis: The social origins of interoceptive inference by Fotopoulou and Tsakiris. *Neuropsychopharmacology*. 19, 29–33. doi: 10.1080/15294145.2017.1295214
- Atzil, S., Gao, W., Fradkin, I., and Barrett, L. F. (2018). Growing a social brain. *Nat. Hum. Behav.* 2, 624–636. doi: 10.1038/s41562-018-0384-6
- Atzil, S., Hendler, T., and Feldman, R. (2011). Specifying the neurobiological basis of human attachment: brain, hormones, and behavior in synchronous and intrusive mothers. *Neuropsychopharmacology*. 36, 2603–2615. doi: 10.1038/npp.2011.172
- Beebe, B., and Lachmann, F. M. (2014). *The Origins of Attachment: Infant Research and Adult Treatment*. New York, NY: Routledge.
- Boston Change Process Study Group. (2005). The “something more” than interpretation revisited: sloppiness and co-creativity in the psychoanalytic encounter. *J. Am. Psychoanal. Assoc.* 53, 693–729. discussion 761–9. doi: 10.1177/00030651050530030401
- Boston Change Process Study Group. (2013). Enactment and the emergence of new relational organization. *J. Am. Psychoanal. Assoc.* 61, 727–749. doi: 10.1177/0003065113496636
- Boston Change Process Study Group. (2018). Moving through and being moved by: embodiment in development and in the therapeutic relationship. *Contemp. Psychoanal.* 54, 299–321. doi: 10.1080/00107530.2018.1456841
- Brockman, R. (2002). Self, object, neurobiology. *Neuropsychopharmacology*. 4, 89–101. doi: 10.1080/15294145.2002.10773382
- Bromberg, P. M. (1996). Standing in the spaces: the multiplicity of self and the psychoanalytic relationship. *Contemp. Psychoanal.* 32, 509–535. doi: 10.1080/00107530.1996.10746334
- Bromberg, P. M. (2014). *Standing in the Spaces: Essays on Clinical Process Trauma and Dissociation*. London: Routledge.
- Carhart-Harris, R. L., Erritzoe, D., Haijen, E., Kaelen, M., and Watts, R. (2018). Psychedelics and connectedness. *Psychopharmacology* 235, 547–550. doi: 10.1007/s00213-017-4701-y
- Carlson, E. A., and Sroufe, L. A. (1995). “Contribution of attachment theory to developmental psychopathology,” in *Developmental Psychopathology*, Vol. 1. Theory and Methods, eds D. Cicchetti and D. J. Cohen (John Wiley & Sons), 581–617.

- Cavicchioli, M., Scalabrini, A., Northoff, G., Mucci, C., Ogliari, A., Maffei, C., et al. (2021). Dissociation and emotion regulation strategies: a meta-analytic review. *J. Psychiatr. Res.* 143, 370–387. doi: 10.1016/j.jpsychires.2021.09.011
- Cieri, F. (2022). Memory for the future: psychodynamic approach to time and self through the default network. *Front. Hum. Neurosci.* 16, 885315. doi: 10.3389/fnhum.2022.885315
- Cieri, F., and Esposito, R. (2019). Psychoanalysis and neuroscience: the bridge between mind and brain. *Front. Psychol.* 10, 1790. doi: 10.3389/fpsyg.2019.01983
- Cieri, F., Zhuang, X., Caldwell, J. Z. K., and Cordes, D. (2021). Brain entropy during aging through a free energy principle approach. *Front. Hum. Neurosci.* 15, 647513. doi: 10.3389/fnhum.2021.647513
- Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Curr. Opin. Neurobiol.* 13, 500–505. doi: 10.1016/S0959-4388(03)00090-4
- Craig, A. D. (2010). The sentient self. *Brain Struct. Funct.* 214, 563–577. doi: 10.1007/s00429-010-0248-y
- Craig, A. D. (2011). Significance of the insula for the evolution of human awareness of feelings from the body. *Ann. N. Y. Acad. Sci.* 1225, 72–82. doi: 10.1111/j.1749-6632.2011.05990.x
- Critchley, H., and Seth, A. (2012). Will studies of macaque insula reveal the neural mechanisms of self-awareness? *Neuron*. 74, 423–426. doi: 10.1016/j.neuron.2012.04.012
- Damasio, A. R. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. New York, NY: Random House.
- Deco, G., Tononi, G., Boly, M., and Kringelbach, M. L. (2015). Rethinking segregation and integration: contributions of whole-brain modelling. *Nat. Rev. Neurosci.* 16, 430–439. doi: 10.1038/nrn3963
- Draijer, N., and Langeland, W. (1999). Childhood trauma and perceived parental dysfunction in the etiology of dissociative symptoms in psychiatric inpatients. *Am. J. Psychiatry* 156, 379–385. doi: 10.1176/ajp.156.3.379
- Ebisch, S. J. H., Scalabrini, A., Northoff, G., Mucci, C., Sergi, M. R., Saggino, A., et al. (2022). S. Intrinsic shapes of empathy: functional brain network topology encodes intersubjective experience and awareness traits. *Brain Sci.* 12, 477. doi: 10.3390/brainsci12040477
- Esposito, R., Cieri, F., di Giannantonio, M., and Tartaro, A. (2018). The role of body image and self-perception in anorexia nervosa: the neuroimaging perspective. *J. Neuropsychol.* 12, 41–52. doi: 10.1111/jnp.12106
- Facco, E., Al Khafaji, B. E., and Tressoldi, P. (2019). In search of the true self. *J. Theor. Philos. Psychol.* 39, 157. doi: 10.1037/teo0000112
- Farina, B., Liotti, M., and Imperatori, C. (2019). The role of attachment trauma and disintegrative pathogenic processes in the traumatic-dissociative dimension. *Front. Psychol.* 10, 933. doi: 10.3389/fpsyg.2019.00933
- Feldman, R. (2007). Parent-infant synchrony and the construction of shared timing: physiological precursors, developmental outcomes, and risk conditions. *J. Child Psychol. Psychiatry* 48, 329–354. doi: 10.1111/j.1469-7610.2006.01701.x
- Feldman, R. (2015). Sensitive periods in human social development: new insights from research on oxytocin, synchrony, and high-risk parenting. *Dev. Psychopathol.* 27, 369–395. doi: 10.1017/S0954579415000048
- Feldman, R., Eidelman, A. I., Sirota, L., and Weller, A. (2002). Comparison of skin-to-skin (kangaroo) and traditional care: parenting outcomes and preterm infant development. *Pediatrics* 110, 16–26. doi: 10.1542/peds.110.1.16
- Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., and Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behav. Dev.* 34, 569–577. doi: 10.1016/j.infbeh.2011.06.008
- Ferenczi, S. (1932a/1988). *The Clinical Diary of Sandor Ferenczi* (ed. J. Dupont, Trans. M. Balint and N. Zarday Jackson). Cambridge, MA: Harvard University Press.
- Ferenczi, S. (1932b/1984). “Confusion of the tongue between adults and the child (the language of tenderness and the language of [sexual] passion). Appendix C,” in: *The Assault on Truth: Freud's Suppression of the Seduction Theory*, ed. J. M. Masson, Trans. J. M. Masson, and I. Loring (New York, NY: Random House), 291–303.
- Filippetti, M. L., Johnson, M. H., Lloyd-Fox, S., Dragovic, D., and Farroni, T. (2013). Body perception in newborns. *Curr. Biol.* 23, 2413–2416. doi: 10.1016/j.cub.2013.10.017
- Fonagy, P., and Bateman, A. W. (2007). Mentalizing and borderline personality disorder. *J. Ment. Health* 16, 83–101.
- Fonagy, P., Gergely, G., and Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *J. Child Psychol. Psychiatry* 48, 288–328. doi: 10.1111/j.1469-7610.2007.01727.x
- Fordham, M. (1957). *New Developments in Analytical Psychology*. London: Routledge & Kegan Paul, 198
- Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychanalysis* 19, 3–28. doi: 10.1080/15294145.2017.1294031
- Freud, S. (1895). *Project for a Scientific Psychology*. London: Hogarth Press.
- Gallagher, I. I. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5
- Gallagher, S. (2005). “Dynamic models of body schematic processes,” in: *Body Image and Body Schema*, eds. H. De Preester, and V. Knockaert (Amsterdam: John Benjamins Publishers), 233–250. doi: 10.1075/aicr.62.15gal
- Gu, X., Hof, P. R., Friston, K. J., and Fan, J. (2013). Anterior insular cortex and emotional awareness. *J. Comp. Neurol.* 521, 3371–3388. doi: 10.1002/cne.23368
- Hofer, M. A. (1994). Hidden regulators in attachment, separation, and loss. *Monogr. Soc. Res. Child Dev.* 59, 192–207. doi: 10.1111/j.1540-5834.1994.tb01285.x
- Kernberg, O. (1967). Borderline personality organization. *J. Am. Psychoanal. Assoc.* 15, 641–685. doi: 10.1177/000306516701500309
- Kernberg, O. F. (1984). *Severe Personality Disorders: Psychotherapeutic Strategies*. New Haven, CT: Yale UP, 227–228.
- Kohut, H. (1971). *The Analysis of the Self*. New York, NY: International Universities Press.
- Levin, B. E. (2006). Metabolic imprinting: critical impact of the perinatal environment on the regulation of energy homeostasis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* 361, 1107–1121. doi: 10.1098/rstb.2006.1851
- Levy, J., Goldstein, A., and Feldman, R. (2017). Perception of social synchrony induces mother-child gamma coupling in the social brain. *Soc. Cogn. Affect. Neurosci.* 12, 1036–1046. doi: 10.1093/scan/nsx032
- Liotti, G. (1992). Disorganized/disoriented attachment in the etiology of the dissociative disorders. *Dissociation* 5, 196–204.
- Llobera, J., Charbonnier, C., Chagué, S., Preissmann, D., Antonietti, J. P., Ansermet, F., et al. (2016). The subjective sensation of synchrony: an experimental study. *PLoS ONE* 11, e0147008. doi: 10.1371/journal.pone.0147008
- Lonstein, J. S., Lévy, F., and Fleming, A. S. (2015). Common and divergent psychobiological mechanisms underlying maternal behaviors in non-human and human mammals. *Horm. Behav.* 73, 156–185. doi: 10.1016/j.yhbeh.2015.06.011
- Lucherini Angeletti, L., Scalabrini, A., Ricca, V., and Northoff, G. (2021). Topography of the anxious self: Abnormal rest-task modulation in social anxiety disorder. *Neuroscientist*. doi: 10.1177/10738584211030497. [Epub ahead of print].
- Lyons-Ruth, K. (2003). Dissociation and the parent-infant dialogue: a longitudinal perspective from attachment research. *J. Am. Psychoanal. Assoc.* 51, 883–911. doi: 10.1177/00030651030510031501
- Lyons-Ruth, K. (2008). Contributions of the mother-infant relationship to dissociative, borderline, and conduct symptoms in young adulthood. *Infant Ment. Health J.* 29, 203–218. doi: 10.1002/imhj.20173
- Main, M., and Hesse, E. (1990). “Parents’ unresolved traumatic experiences are related to infant disorganized attachment status: is frightened and/or frightening parental behavior the linking mechanism?,” in: *Attachment in the Preschool Years*, eds. M. C. Greenberg, D. Cicchetti, and E. M. Cummings (Chicago, IL: University of Chicago Press), 161–182.
- Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0
- Mucci, C. (2013). *Beyond Individual and Collective Trauma. Intergenerational Transmission, Psychoanalytic Treatment, and the Dynamics of Forgiveness*. London: Karnac Books.
- Mucci, C. (2014). Trauma, healing and the reconstruction of truth. *Am. J. Psychoanal.* 74, 31–47. doi: 10.1057/ajp.2013.37
- Mucci, C. (2017). Ferenczi’s Revolutionary Therapeutic Approach. *Am. J. Psychoanal.* 77, 239–254. doi: 10.1057/s11231-017-9104-7
- Mucci, C. (2018a). *Borderline Bodies: Affect Regulation Therapy for Personality Disorders (Norton Series on Interpersonal Neurobiology)*. New York, NY: WW Norton & Company.
- Mucci, C. (2018b). “Implicit memory, unexpressed unconscious, and trauma theory: the turn of the screw between contemporary psychoanalysis and neuroscience,” in: *Unexpressed Unconscious, Implicit Memory, and Clinical Work* (London: Routledge), 99–129. doi: 10.4324/9780429484629-5
- Mucci, C. (2019). Traumatization through human agency: “embodied witnessing” is essential in the treatment of survivors. *Am. J. Psychoanal.* 79, 540–554. doi: 10.1057/s11231-019-09225-y

- Mucci, C. (2021a). "Intersubjectivity and psychopathology: borderline and psychosomatic bodies "at the mind's limits"," in: *International Forum of Psychoanalysis*, Vol. 30, No. 3 (London: Routledge), 156–166. doi: 10.1080/0803706X.2021.1950930
- Mucci, C. (2021b). Dissociation vs repression: a new neuropsychanalytic model for psychopathology. *Am. J. Psychoanal.* 81, 82–111. doi: 10.1057/s11231-021-09279-x
- Mucci, C. (2022). *Resilience and Survival: Understanding and Healing Intergenerational Trauma*. London: Karnac Books.
- Mucci, C., and Scalabrini, A. (2021). Traumatic effects beyond diagnosis: The impact of dissociation on the mind-body-brain system. *Psychoanal. Psychol.* 38, 279–289. doi: 10.1037/pap0000332
- Nakata, T., and Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant. Behav.* 27, 455–464. doi: 10.1016/j.infbeh.2004.03.002
- Nelson, B., Parnas, J., and Sass, L. A. (2014). Disturbance of minimal self (ipseity) in schizophrenia: clarification and current status. *Schizophr. Bull.* 40, 479–482. doi: 10.1093/schbul/sbu034
- Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. *Cogn. Neurosci.* 7, 203–222. doi: 10.1080/17588928.2015.1111868
- Northoff, G. (2018). The brain's spontaneous activity and its psychopathological symptoms - "Spatiotemporal binding and integration. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 80(Pt B):81–90. doi: 10.1016/j.pnpbp.2017.03.019
- Northoff, G., and Bermpohl, F. (2004). Cortical midline structures and the self. *Trends Cogn. Sci.* 8, 102–107. doi: 10.1016/j.tics.2004.01.004
- Northoff, G., and Huang, Z. (2017). How do the brain's time and space mediate consciousness and its different dimensions? Temporo-spatial theory of consciousness (TTC). *Neurosci. Biobehav. Rev.* 80, 630–645. doi: 10.1016/j.neubiorev.2017.07.013
- Northoff, G., and Panksepp, J. (2008). The trans-species concept of self and the subcortical-cortical midline system. *Trends Cogn. Sci.* 12, 259–264. doi: 10.1016/j.tics.2008.04.007
- Northoff, G., Qin, P., and Feinberg, T. E. (2011). Brain imaging of the self-conceptual, anatomical and methodological issues. *Conscious. Cogn.* 20, 52–63. doi: 10.1016/j.concog.2010.09.011
- Northoff, G., and Scalabrini, A. (2021). "Project for a spatiotemporal neuroscience" - brain and psyche share their topography and dynamic. *Front. Psychol.* 12, 717402. doi: 10.3389/fpsyg.2021.717402
- Northoff, G., and Smith, D. (2022). The subjectivity of self and its ontology: from the world-brain relation to the point of view in the world. *Theory Psychol.* doi: 10.1177/09593543221080120
- Northoff, G., Vatansever, D., Scalabrini, A., and Stamatakis, E. A. (2022). Ongoing brain activity and its 976 role in cognition: dual versus baseline models. *Neuroscientist.* doi: 10.1177/10738584221081752
- Northoff, G., Wainio-Theberge, S., and Evers, K. (2020a). Is temporo-spatial dynamics the "common currency" of brain and mind? In quest of "spatiotemporal neuroscience". *Phys. Life Rev.* 33, 34–54. doi: 10.1016/j.plrev.2019.05.002
- Northoff, G., Wainio-Theberge, S., and Evers, K. (2020b). Spatiotemporal neuroscience - what is it and why we need it. *Phys. Life Rev.* 33, 78–87. doi: 10.1016/j.plrev.2020.06.005
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.
- Panksepp, J., and Biven, L. (2012). *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions (Norton Series on Interpersonal Neurobiology)*. New York, NY: WW Norton & Company.
- Park, H. D., Correia, S., Ducorps, A., and Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nat. Neurosci.* 17, 612–618. doi: 10.1038/nn.3671
- Park, H. D., and Tallon-Baudry, C. (2014). The neural subjective frame: from bodily signals to perceptual consciousness. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 369, 20130208. doi: 10.1098/rstb.2013.0208
- Porges, S. (2011). *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, Self-Regulation*. New York, NY: Norton.
- Qin, P., Wang, M., and Northoff, G. (2020). Linking bodily, environmental and mental states in the self-A three-level model based on a meta-analysis. *Neurosci. Biobehav. Rev.* 115, 77–95. doi: 10.1016/j.neubiorev.2020.05.004
- Riem, M. M., Bakermans-Kranenburg, M. J., Pieper, S., Tops, M., Boksem, M. A., Vermeiren, R. R., et al. (2011). Oxytocin modulates amygdala, insula, and inferior frontal gyrus responses to infant crying: a randomized controlled trial. *Biol. Psychiatry.* 70, 291–297. doi: 10.1016/j.biopsych.2011.02.006
- Scalabrini, A., Ebisch, S. J. H., Huang, Z., Di Plinio, S., Perrucci, M. G., Romani, G. L., et al. (2019). Spontaneous brain activity predicts task-evoked activity during animate versus inanimate touch. *Cereb. Cortex.* 29, 4628–4645. doi: 10.1093/cercor/bhy340
- Scalabrini, A., Esposito, R., and Mucci, C. (2021a). Dreaming the unrepressed unconscious and beyond: repression vs dissociation in the onerific functioning of severe patients. *Res Psychother.* 24, 545. doi: 10.4081/ripppo.2021.545
- Scalabrini, A., Huang, Z., Mucci, C., Perrucci, M. G., Ferretti, A., Fossati, A., et al. (2017). How spontaneous brain activity and narcissistic features shape social interaction. *Sci. Rep.* 7, 9986. doi: 10.1038/s41598-017-10389-9
- Scalabrini, A., Mucci, C., Angeletti, L. L., and Northoff, G. (2020b). The self and its world: a neuro-ecological and temporo-spatial account of existential fear. *Clin. Neuropsychiatry* 17, 46–58.
- Scalabrini, A., Mucci, C., Esposito, R., Damiani, S., and Northoff, G. (2020a). Dissociation as a disorder of integration - On the footsteps of Pierre Janet. *Prog. Neuropsychopharmacol. Biol. Psychiatry.* 101, 109928. doi: 10.1016/j.pnpbp.2020.109928
- Scalabrini, A., Mucci, C., and Northoff, G. (2018). Is our self related to personality? a neuropsychodynamic model. *Front. Hum. Neurosci.* 12, 346. doi: 10.3389/fnhum.2018.00346
- Scalabrini, A., Schimmenti, A., De Amicis, M., Porcelli, P., Benedetti, F., Mucci, C., et al. (2022). The self and its internal thought: in search for a psychological baseline. *Conscious. Cogn.* 97, 103244. doi: 10.1016/j.concog.2021.103244
- Scalabrini, A., Vai, B., Poletti, S., Damiani, S., Mucci, C., Colombo, C., et al. (2020c). All roads lead to the default-mode network-global source of DMN abnormalities in major depressive disorder. *Neuropsychopharmacology* 45, 2058–2069. doi: 10.1038/s41386-020-0785-x
- Scalabrini, A., Wolman, A., and Northoff, G. (2021b). The self and its right insula-differential topography and dynamic of right vs. left insula. *Brain Sci.* 11, 1312. doi: 10.3390/brainsci11101312
- Scalabrini, A., Xu, J., and Northoff, G. (2021c). What COVID-19 tells us about the self: the deep intersubjective and cultural layers of our brain. *Psychiatry Clin. Neurosci.* 75, 37–45. doi: 10.1111/pcn.13185
- Schimmenti, A., and Caretti, V. (2016). Linking the overwhelming with the unbearable: developmental trauma, dissociation, and the disconnected self. *Psychoanal. Psychol.* 33, 106. doi: 10.1037/a0038019
- Schimmenti, A., and Sar, V. A. (2019). correlation network analysis of dissociative experiences. *J. Trauma Dissociation.* 20, 402–419. doi: 10.1080/15299732.2019.1572045
- Schore, A. N. (2000). Attachment and the regulation of the right brain. *Attach. Hum. Dev.* 2, 23–47. doi: 10.1080/146167300361309
- Schore, A. N. (2001a). The effects of a secure attachment relationship on right brain development, affect regulation, and infant mental health. *Infant Ment. Health J.* 22, 7–66. doi: 10.1002/1097-0355(200101/04)22:1<7::AID-IMHJ2>3.0.CO;2-N
- Schore, A. N. (2001b). The effects of early relational trauma on right brain development, affect regulation, and infant mental health. *Infant Ment. Health J.* 22, 201–269. doi: 10.1002/1097-0355(200101/04)22:1<201::AID-IMHJ8>3.0.CO;2-9
- Schore, A. N. (2003a). *Affect Dysregulation and Disorders of The Self (Norton Series on Interpersonal Neurobiology)*. New York, NY: WW Norton & Company.
- Schore, A. N. (2003b). *Affect Regulation and the Repair of the Self (Norton Series on Interpersonal Neurobiology)*, Vol. 2. New York, NY: WW Norton & Company.
- Schore, A. N. (2005). Attachment, affect regulation, and the developing right brain: Linking developmental neuroscience to pediatrics. *Pediatr. Rev.* 26, 204–217. doi: 10.1542/pir.26-6-204
- Schore, A. N. (2009). Relational trauma and the developing right brain: an interface of psychoanalytic self psychology and neuroscience. *Ann. N. Y. Acad. Sci.* 1159, 189–203. doi: 10.1111/j.1749-6632.2009.04474.x
- Schore, A. N. (2011). Foreword. In P. Bromberg: the shadow of the tsunami and the growth of the relational mind (pp. ix–xxxvii). New York: Routledge.
- Schore, A. N. (2012). *The Science of the Art of Psychotherapy (Norton Series on Interpersonal Neurobiology)*. New York, NY: Norton.
- Schore, A. N. (2021). The interpersonal neurobiology of intersubjectivity. *Front. Psychol.* 12, 648616. doi: 10.3389/fpsyg.2021.648616
- Seth, A. K., and Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 371, 20160007. doi: 10.1098/rstb.2016.0007

- Siegel, D. J. (1999). *The Developing Mind: Toward a Neurobiology of Interpersonal Experience*. New York, NY: Guilford Press.
- Solms, M. (2020). New project for a scientific psychology: general scheme. *Neuropsychanalysis* 22, 5–35. doi: 10.1080/15294145.2020.1833361
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. Profile books.
- Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238.
- Solms, M., and Panksepp, J. (2012). The “id” knows more than the “ego” admits: neuropsychanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sci.* 2, 147–175. doi: 10.3390/brainsci2020147
- Spagnolo, R., and Northoff, G. (2021). *The Dynamic Self in Psychoanalysis. Neuroscientific Foundations and Clinical Cases*. New York, NY: Routledge.
- Stern, D. N. (1985). *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. London: Karnac Books.
- Stern, D. N. (2004). *The Present Moment in Psychotherapy and Everyday Life (Norton Series on Interpersonal Neurobiology)*. New York, NY: WW Norton & Company.
- Strachey, J. (1961). *Editor's Introduction. Sigmund Freud Civilization and Its Discontents*. Berlin, 3–11.
- Tallon-Baudry, C., Campana, F., Park, H. D., and Babo-Rebelo, M. (2018). The neural monitoring of visceral inputs, rather than attention, accounts for first-person perspective in conscious vision. *Cortex.* 102, 139–149. doi: 10.1016/j.cortex.2017.05.019
- Trevarthen, C. (1993). “The Self Born in Intersubjectivity: the psychology of an Infant Communicating,” in *The Perceived Self: Ecological and Interpersonal Sources of Self-knowledge*, ed. U. Neisser (New York, NY: Cambridge University Press), 121–173. doi: 10.1017/CBO9780511664007.009
- Trevarthen, C., and Aitken, K. J. (2001). Infant intersubjectivity: research, theory, and clinical applications. *J. Child Psychol. Psychiatry.* 42, 3–48. doi: 10.1111/1469-7610.00701
- Tsakiris, M. (2017). The multisensory basis of the self: from body to identity to others. *Q. J. Exp. Psychol.* 70, 597–609. doi: 10.1080/17470218.2016.1181768
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25. doi: 10.1146/annurev.psych.53.100901.135114
- Ulmer Yaniv, A., Salomon, R., Waidergoren, S., Shimon-Raz, O., Djalovski, A., Feldman, R., et al. (2021). Synchronous caregiving from birth to adulthood tunes humans' social brain. *Proc. Natl. Acad. Sci. USA.* 118, e2012900118. doi: 10.1073/pnas.2012900118
- Watts, R., Day, C., Krzanowski, J., Nutt, D., and Carhart-Harris, R. (2017). Patients' accounts of increased “connectedness” and “acceptance” after psilocybin for treatment-resistant depression. *J. Humanist. Psychol.* 57, 520–564. doi: 10.1177/0022167817709585
- Winnicott, D. W. (1965). *The Family and Individual Development*. Basic Books, Inc.



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Andrea Scalabrini,
University of Bergamo, Italy
Chiara Pazzagli,
University of Perugia, Italy

*CORRESPONDENCE

Riccardo Williams
riccardo.williams@uniroma1.it

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 29 June 2022

ACCEPTED 10 August 2022

PUBLISHED 08 September 2022

CITATION

Williams R and Trentini C (2022) Two
modes of being together: The levels
of intersubjectivity and human
relatedness in neuroscience
and psychoanalytic thinking.
Front. Hum. Neurosci. 16:981366.
doi: 10.3389/fnhum.2022.981366

COPYRIGHT

© 2022 Williams and Trentini. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Two modes of being together: The levels of intersubjectivity and human relatedness in neuroscience and psychoanalytic thinking

Riccardo Williams* and Cristina Trentini

Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome,
Rome, Italy

The notion of intersubjectivity has achieved a primary status in contemporary psychoanalytic debate, stimulating new theoretical proposals as well as controversies. This paper presents an overview of the main contributions on inter-subjectivity in the field of neurosciences. In humans as well as—probably—in other species, the ability for emotional resonance is guaranteed early in development. Based on this capacity, a primary sense of connectedness is established that can be defined inter-subjective in that it entails sharing affective states and intentions with caregivers. We propose to define such a form of inter-subjectivity as *contingent*, since the infant's early abilities for resonance do not imply the more generalized capacity to permanently conceive of the relationship outside the realm of current interactions and the infant-caregiver's mutual correspondence of internal states. This form of connection, hence, results in a self-referential, bodily, and affectively codified, context- and time dependent, like-me experience of interactions. The gradual maturation of brain structures and processes as well as interactive experiences allow proper intersubjectivity exchanges, grounded on new intentional and representational capacities, to evolve. In this more mature form of intersubjectivity, the individual is allowed to conceive of her own psychic space both as distinct and as possibly connected with the other's contents and experience, even in the absence of current behavioral indicators of such correspondence. This multi-layered model of intersubjectivity, which

is embraced by current neuroscience research, seems to allow for new interpretations of psychoanalytic models of human relatedness based upon classic clinical observations.

KEYWORDS

contingent intersubjectivity, mentalization, self-other representations, psychoanalysis, neuroscience, mirror neurons, affective resonance

Introduction

The recent main theoretical and clinical innovations in psychoanalysis stem from a view that is based on the notion of primary intersubjectivity (Stern, 2005). The psychoanalytic bi-personal model conceives of the psychic as the product of the interiorization of real early interpersonal exchanges and human motivation for relatedness (Eagle, 2011). This new perspective has also led to previous psychoanalytic clinical theories models—founded on developmental notions, such as the phase of primary narcissism (Freud, 1914), normal autism and early symbiosis (Mahler et al., 1975), primary self-other undifferentiation (Winnicott, 1945; Jacobson, 1964; Kohut, 1971), and primacy of phylogenetically inborn phantasies that shape early representations of interactions (Klein, 1946; Kernberg, 1985)—to be questioned.

The current intersubjective or relational perspective heavily relies on infant research evidence (Beebe and Lachmann, 2002), claiming to abandon the so-called adult-morphic and pathomorphic psychoanalytical reconstructions of early infant development (Peterfreund, 1978). However, among developmental researchers, a consensus regarding the true intersubjective nature of early interactive experiences (Carruthers and Smith, 1996) remains far from being reached. Two main perspectives struggle in this area. According to the simulationist point of view, the early capacity to understand the other's intentions and experience sharing is warranted by early mechanisms of imitation that allow the other's intentions to be internally reproduced and matched with the observer's own experience and intentions (Goldman, 2006). In the alternative view (the mentalistic perspective), no experience of sharing intentions can be achieved before the capacity for attributing mental states to the self and the other is established (Tomasello, 1999).

In this paper, we scrutinize the contributions of neuroscience in modeling the experience of interactions during development. To discuss the data from this field, we consider some basic prerequisites for the recognition of intersubjective capacity that are usually considered in developmental research, such as the infants' ability to: a) form self-other unified representations; b) understand the intentions pertaining to self and others' behaviors; c) establish a self-other

differentiation; and d) be aware and understand that their own behaviors and those of others have the same intentions.

A neuroscience perspective of intersubjectivity

The bodily multimodal nature of early self and other's representations

The bodily multimodal nature of early self and other's representations has been widely considered in neuroscience literature. A recent perspective comes from Atzil et al. (2018), who propose that representations of social objects are built upon the regular association between interoceptive information about allostasis and exteroceptive information deriving from exchanges with the caregivers. Such a perspective introduces a more socially-oriented interpretation of the original definition of "allostasis," according to which the brain is constantly engaged in regulating the organism's internal milieu by anticipating needs and preparing to fulfill them before they arise (Sterling, 2012; Sterling and Laughlin, 2015). Atzil et al. (2018) claim that there is no "core social knowledge" at birth (Carey and Spelke, 1996; Spelke and Kinzler, 2007): this form of knowledge rather arises from the gradual learning about social agents and social behaviors. Since the brain "categorizes sensory information to predict about allostasis (...), sensory regularities (such as a face) will [thus] become concepts more rapidly if they impact allostasis. The association between allostasis and a human agent [such as the mother] will result in learning an important social concept: 'mommy'" (Atzil et al., 2018; p. 630). This (in our opinion, constructivist) model postulates that rudimentary social concepts develop in early infancy in the form of multimodal representations (such as maternal face) and become more abstract with development. Moreover, this model postulates that, through social regulation of allostasis, infants also acquire social competencies—such as synchrony—that they learn to intentionally use to regulate both their own and others' allostasis (Atzil et al., 2011, 2018).

Indeed, supporters of primary intersubjectivity also believe that the capacity to build representations of the self and the other

are based on an intermodal code. At the same time—assuming a very different perspective from that presented above—, they also believe that, already at birth, infants are aware of the presence of the other persons and (by virtue of that) can build representations of self-other interactions in the domain of actual experiences with others (Trevvarthen, 1974, 1979, 1993, 1998).

The initial behavioral indicator of early intersubjectivity is represented by infants' ability to imitate facial movements performed by an adult (such as the opening of the mouth and the protrusion of the tongue or lips), as documented by Meltzoff and Moore (1997) in their laboratory study on six newborns (of whom one was only 60 min old). Research has proved that imitation is not an automatic and involuntary reflex-like phenomenon, but a behavior toward which newborns are strongly motivated. Newborns, not only imitated gestures, but they also spontaneously “provoke” previously imitated gestures, waiting for the other to respond (Nagy and Molnar, 2004). Moreover, newborns can correct their own movements to make them converge with those of the observed adult (Meltzoff and Moore, 1997), and can reproduce gestures after a 24-h delay, from at least 6 weeks of age (Meltzoff and Moore, 1992, 1994). In this perspective imitative behaviors are interpreted as evidence of a precocious sense of self as differentiated from others and agent in the environment (Rochat and Striano, 2000). As Metzoff and Moore (1977) have stated, during imitation, the infant compares “the sensory information from his own unseen motor behavior to a ‘supramodal’ representation of the visually perceived gesture and construct the match required (...). [Imitative behaviors] are (...) accomplished through an active matching process and mediated by an abstract representational system” (p. 78). This “active intermodal mapping” (Meltzoff and Moore, 1997) is possible because the perception and production of self and others' actions are represented within a common framework. Hence, in this stage, the infant “feels *what* the other feels.”

Numerous studies have provided the evidence that imitative capability is ensured by neural mirror mechanisms, allowing a shared mapping (and thus a common framework) between self and others, at the bodily level. Mirror neurons (MNs)—discovered in the ventral premotor cortex (vPMC) and the inferior parietal lobe (IPL) of macaque monkeys (Gallese et al., 1996; Rizzolatti et al., 1996)—are a distinct class of neurons that discharge when an individual performs a goal-directed action or observes someone else performing the same action. MNs—and the non-conscious, prereflective, and presymbolic functional mechanism that they underpin [i.e., the *embodied simulation* (Gallese, 2003)]—allow individuals who are confronting others' behaviors to experience a specific phenomenal state of “intentional attunement” (Gallese, 2006). Such a condition generates a peculiar quality of familiarity with other individuals that is produced by

the resonance of their emotions and intentions with the observer's simulation.

Recent electroencephalography (EEG) studies have shown that shared representations between self and others are reflected by mirror mechanisms in the infant sensorimotor cortex (Simpson et al., 2014; Southgate et al., 2009, 2010), similar to those that are found in adult brains (Gallese, 2014). Notably, positron emission tomography (PET) studies have documented that metabolic activity is highest in the sensorimotor cortex, already before 5 weeks of age (Chugani, 1994; Chugani et al., 1987).

EEG studies on mirror mechanisms in infants have focused on the sensorimotor alpha (or mu) rhythm during action observation and action execution (Cuevas et al., 2014; Marshall et al., 2002; Marshall and Meltzoff, 2011). The mu rhythm is an EEG oscillation in the alpha frequency range (recorded over the central electrodes) that is generated in the resting state and desynchronized (i.e., attenuated or suppressed) prior to or during motor events. In infants, the observation of an experimenter who is performing a goal-directed action using a particular body part (hands or feet) is associated with desynchronization of the mu rhythm of the corresponding area of the body in the infant sensorimotor cortex (Marshall and Meltzoff, 2014; Saby et al., 2013). Similar somatotopic patterns have also been observed when infants perform the same actions as an experimenter (Marshall et al., 2013). Such somatotopic organization is considered as an index of the “intercorporeal mapping of corresponding body parts between self and other” (Marshall et al., 2013; p. 22), which allows an infant to engage in early imitation (Meltzoff, 1988; Meltzoff and Moore, 1997). It may be suggested that sensorimotor cortex guarantees the “supramodal” mechanisms (Meltzoff and Moore, 1997) that allow the infant to share the sensory feedbacks coming from the caregiver's behaviors and those evoked by his own movements (Meltzoff and Marshall, 2018). Furthermore, data somatotopic organization may be considered by supporters of primary intersubjectivity as neuroscientific evidence of the “like me” simulationist framework (Meltzoff, 2007, 2013), according to which infants can parse a similarity (or equivalence) between their own bodily acts and those of others.

In a critical approach to the perspective of primary intersubjectivity, whether imitative behaviors clearly reflect the primary capacity to *understand* another's intentions is questioned. Similar forms of imitations exist in other species (such as macaques), in which forms of proper intersubjective exchanges are not otherwise displayed (Gallese, 2014; Tomasello, 1999). Another critical point concerns the nature of the experiences that are ensured by early imitative behaviors. This issue has been discussed in the neuroscience literature, demonstrating that inner reproduction of another's observed behavior is not followed by an *aware experience* of one's own intention or that of others (Avenanti et al., 2005; Caramazza et al., 2014).

According to these criticisms, the low-level processing of observed actions in imitative behaviors does not suffice for an experience of intersubjective sharing which, instead, requires higher-level cognitive processes that allow intentional states to be inferred (Heyes and Catmur, 2022). As Di Bernardo (2021) has stated, “in order to establish full and effective emotional communication between two people, it is necessary for each of the people involved to let their own state of mind be influenced by that of the other, so that they ‘feel’ it and tune in to it” (p. 8). In response to these criticisms, simulationist theorists have focused on observational and neuroscience studies that appear to support the early achievement of intersubjective capacities.

Understanding intentionality

After early imitations, other routines of mutually oriented interactions emerge gradually, including protoconversations (Trevarthen, 1998), affective tuning (Stern, 1985), and turn-taking behaviors (Weinberg and Tronick, 1997).

Neuroscience research has evidenced that these interactions are regulated by the right temporoparietal junction (rTPJ) (for a review, see Schore, 2021), an area located at the intersection of the posterior end of the superior temporal sulcus (STS), the inferior parietal cortex (IPC), and the lateral occipital cortex (LOC). Coherently, near-infrared spectroscopy (NIRS) studies on young infants have documented that activity in rTPJ is enhanced in response to social signals (such as human voice; Grossmann et al., 2010) and reciprocal interactions (Hakuno et al., 2020). Notably, the right lateralization of this cerebral system is consistent with Schore’s (2021) perspective on interpersonal neurobiology of intersubjectivity and with what Decety and Chaminade (2003) have reported in the conclusions of their review: “intersubjective processes are largely dependent upon the right hemisphere resources, which are the first to develop” (p. 591).

In these types of interactions, infants show the capacity to modulate their behaviors by anticipating another’s intention, to reach a condition of sharing their affective experience (Stern, 1985). The experience of intersubjective sharing necessarily implies the capacity to connect one’s own intentions to the other’s intentions: therefore, it has as its prerequisite the capacity to understand intentions.

Current empirical and theoretical orientations conceive intentionality not as a unitary capacity, but as a hierarchy of abilities (Bekoff, 2007; Dennett, 1991; Tomasello and Call, 1997). At the first level (which is shared by many species), intentionality implies the capacity to have a belief about an object. The second level of intentionality entails the capacity to have a belief about another agent’s belief and regulate one’s behavior consequently. Infant observation has provided

evidence for the presence early in infancy of this level of intentionality and, by virtue of that, of primary forms of intersubjective sharing. This second level does not require the capacity to explicitly represent beliefs as such and can be observed in many species, including animals’ deceitful and playful behaviors (Tomasello et al., 2003; Jamieson and Bekoff, 1996). The third level of intentionality (which is instead most typical of human beings) is characterized by a reflective form of intentional attributions, in which an individual is able not only to have a belief about another individual’s belief but also to see whether the other’s belief corresponds or not to actual reality (Dennett, 1991). This higher form of intentional stance is linked to a more basic capacity to explicitly attribute and reflect upon mental states (Tomasello, 1999).

MNs are considered to provide a neuroscientifically viable account of intentional understanding (Gallese, 2014). The distinctive characteristic of MNs is that observing a motor act activates the same motor cerebral areas that are required by the observer to execute that same action (Rizzolatti and Sinigaglia, 2008). In line with this, the observer can directly understand others’ actions (the *what*) and to ascribe them intentions (the *why*), without the mediation of any cognitive or inferential processes (Rizzolatti and Sinigaglia, 2007; Rizzolatti and Fogassi, 2014). In recent years, the MNs perspective has proposed an innovative interpretation of intention understanding and early forms of sharing. As stated by Gallese (2014), the intersubjective connection that is allowed by MNs at early stage of development remains at a “subpersonal” level and do not fully account for the complexity of interpersonal experiences that occur in later stages of development, although they have to be regarded as the building blocks of any more complex intersubjective experience in human species. It has been stressed that the understanding and sharing of intentions is codified “via a mechanism of action representation [in which the insula plays a fundamental role] shaping emotional content, such that we ground our empathic resonance in the experience of our acting body and the emotions associated with specific movements” (Carr et al., 2003; p. 5502). It follows that this type of intersubjective connection is limited to the sensory-motor repertoire that is available to the observer (Buccino et al., 2004; Simpson et al., 2014)—i.e., the infant’s activation of single schemes of movements—, and that does not fully replicate the complexity of a caregiver’s behaviors.

Further, the experience of “being together” (Stern, 2004) is circumscribed to the duration of the ongoing interaction with the caregiver and is conceived as occurring at a subpersonal, pre-reflective level, where a sense of being a self in the relationship with another self is not yet developed. Given these features of the experience of early intentional understanding, we propose defining this first level of sharing as “contingent intersubjectivity.”

In the current debate on primary intersubjectivity, there is another key controversy to be dealt with.

Self-other differentiation

For of the experience of *sharing* to occur, an infant must be able to achieve a sense of “ownership” of personal experience that allows her to differentiate her own from other’s experience (Gallagher, 2000): and thus, “to feel *that* the other feels.” Without this capacity, the sharing of experience warranted by the MNs simulation model has to be meant as a source of emotional contagion (de Waal, 1996). To understand the birth of the ability to create the experience of ownership and the self-other differentiation, current neuroscience research has proposed a cogent model of the development of the self.

Recent contributions conceive the self as a multilayered construct which originates in the course of the development, drawing on different sources of internal and external information, and also recruiting diverse brain structure into a progressively integrated complex neural network (Northoff and Scalabrini, 2021; Qin et al., 2020). Assuming this perspective, the large-scale meta-analysis on neuroimaging studies by Qin et al. (2020) proposes a nested hierarchy model of self which includes three intimately connected layers of processing: interoceptive-processing, exteroceptive-processing, and mental-self-processing. This model implies that cerebral regions of the lower level are included in the next higher level and then implemented with additional regions. Therefore, each of the hierarchical layers of self-processing recruits both overlapping and separate brain areas.

At a first level, the implicit (that is, not involving attention) interoceptive-processing is generated by the multimodal representation of the signals provided by the activation of the cardio-circulatory, respiratory, gastrointestinal, and urogenital apparatuses (Fotopoulou and Tsakiris, 2017; Tsakiris, 2017). The integration of such interoceptive signals is mediated by several brain areas, including the bilateral insula, dorsal anterior cingulate cortex (dACC), thalamus, and bilateral parahippocampus gyrus. This network is similar across all mammalian species and comprises mainly regions belonging to the salience network (Menon, 2011). The multimodal representation provided by the integration of interoceptive stimuli may be considered as the somatic marker that establishes the *proto-self* (Damasio, 2010), without which the more advanced aspects of the self could not take place (Gallagher, 2000).

The sense of the self—conceived of as both differentiated and in relation with other individuals—relies on the capacity to attribute the experience derived from internal processes (that is interoceptive, motor, emotional, and cognitive processes) to an objectified representation of the self. The sense of *ownership* characterizing the basic as well as the fully-developed (that is, reflective) forms of self-awareness is based upon this process of self-attribution. For many years, it has been proposed that the first form of objective self-representation was reached at a visual level and implied the capacity for the infant to recognize herself

in the mirror, an ability appearing not before 16–18 months of age (Gallup, 1998). More recently, it has been suggested that a process of objective self-recognition can be operated earlier, drawing on other sources of exteroceptive information, such as visual, tactile, proprioceptive, and multi-sensory information (Gallagher, 2005). Assuming this perspective, the objective self-recognition gradually emerges in the course of development, involving (at early developmental stages) only singular aspects of the body scheme (e.g., face, hands) and motor programs (Fotopoulou and Tsakiris, 2017). In line with the model proposed by Atzil et al. (2018), the possibility to match interoceptive with exteroceptive information is thought to be allowed by the existence of Bayesian computational mechanisms that predict the frequency of the co-occurrence in the activation of these two types of information (Tsakiris, 2017).

In the hierarchy model proposed by Qin et al. (2020), the second layer (that is the exteroceptive-processing) not only integrates exteroceptive, proprioceptive, and multisensory proprioceptive inputs, but it also links the intero- and exteroceptive body signals with signals coming from external environment, modulating basic self-other boundaries (Park and Blanke, 2019; Suzuki et al., 2013; Tsakiris, 2017). These stimuli are processed by a cerebral network that includes the bilateral insula, anteromedial prefrontal cortex (amPFC), PMC, and bilateral TPJ. This layer also comprises regions typically involved in face-recognition (such as the right fusiform gyrus) and sensorimotor areas (such as the postcentral gyrus). Interestingly, Northoff and Scalabrini (2021) have suggested that “as these regions process inputs from different sensory modalities, they may be key in not only integrating extero- and proprioceptive modalities but also different exteroceptive sensory modalities, that is, cross-modal integration” (p. 7).

The third and highest level (that is the mental-self-processing) includes into the contents of the self also stimuli generated by higher order cognitive processes (rather than only body-based physical) stimuli (Qin et al., 2020). This layer recruits the bilateral insula, PMC, bilateral TPJ, and also crucial areas of the default-mode network (DMN)—such as the pregenual anterior cingulate cortex (pACC)/amPFC, posterior cingulate cortex (PCC)—that are implicated in self-referential processes.

Taken together, these data seem to show that any level of self-representation always implies the activation of those brain structures that are recruited at the level of the core interoceptive self-representation. This evidence has important consequences in that it shows that the contents of self-representations are unavoidably shaped by bodily experiences connected to physical needs and emotional activations, as otherwise stated by the psychoanalytic theories. Indeed, the activation of the insula may represent the somatic marker that enables to distinguish between what belongs to the self (because associated to bodily interoceptive information) from what has to be regarded as non-me. TPJ is implicated in high-order cognitive functions—such

as theory of mind (TOM; Chan and Lavalée, 2015; Schurz et al., 2017) and perspective taking (Schurz et al., 2013)—that require a concomitant representation of both self and other (Schurz et al., 2013). By virtue of their anatomical connections (Saur et al., 2008), the TPJ and insula are thought to collaborate to provide the bodily self-consciousness (Park and Blanke, 2019), which “could serve as the basis for further co-representation of social information pertaining to both self and other, to enable efficient interactions with the external world” (Qin et al., 2020; p. 89).

Other authors have specifically investigated the question of self-other differentiation with regard to the domain of activation of MNs. Results from this research area seem to show a degree of convergence with the nested hierarchy model of self, proposed by Qin et al. (2020).

Several studies have mapped the activation of specific cerebral areas while subjects were executing, thinking of, or simply imagining a plan of action in the first or third person, finding all the experimental conditions activate areas with MNs. Moreover, when subjects were asked to observe or think of another agent’s finalized behavior without assuming the third-person perspective, the task automatically led to a default mode (DM) of self-attribution of the feelings, that were produced by the simulation (Decety and Ickes, 2009). The DM of self-attribution is associated with activation of the mPFC, pACC, PCC, and temporopolar cortex (TPC) (Davey et al., 2016; Dixon et al., 2022; Northoff et al., 2006); moreover, it is also associated with areas involved in the synthesis of proprioceptive inputs (such as the insula) that modulate the integration of the body scheme (Cabanis et al., 2013).

The DMN rules out the possibility of proper intersubjective sharing and is used to account for the experience of *emotional contagion* (de Waal, 1996), which has been observed in certain clinical conditions (such as autism and schizophrenia) following imitative behaviors (Singer and Frith, 2005). Conversely, tasks that involve imitation or attribution of intentions to another agent are linked to activation of the right inferior parietal lobule (IPL), which appears to mediate the basic distinction between actions that are generated by the self and those by another agent (Decety and Jackson, 2004).

According to the neuroscience literature (Decety and Ickes, 2009; Gallese, 2014), the activation of mirror mechanisms during imitation is accompanied by the capacity to distinguish the sources of internal and external agency as early as 3 months of age (Gallese, 2014). As discussed above, the possibility to establish a sense of sharing necessarily requires a self-other differentiation (not to be experienced as an affective contagion). This pre-condition relies on the sense of ownership that is carried out through the coupling of interoceptive experience to an at least partially integrated exteroceptive representation of the self (including proprioceptive feedback involved in the sense of personal agency). It still remains matter of debate and empirical research whether this objectified integrated self-representation can initially be brought about by the mere

capacity to distinguish the sources of agency emerging at this stage (Marraffa and Meini, 2019; Gallese, 2014).

Also, mentalistic intersubjective exchange of later stages of development hinges on a sense of ownership of one’s intentional states, conveyed by a basic self-monitoring mechanism of agency and bodily activation (Cermolacce et al., 2007). This assumption is supported on several levels. The activation of specific cerebral areas that govern the attribution of the sources of an action to the self or to the other has been observed in tasks that involve direct imitation and in the imagining plans of action, thoughts and emotions, autobiographical memory, and attribution of personal pronouns (Decety and Ickes, 2009), and is reported to be inverted in certain subjects (such as those with schizophrenia), who show a deficit in holding a sense of ownership of their own actions and mental processes (Farrer et al., 2003).

Understanding the other and the self as mental agents

As Fonagy et al. (2002) have argued, in everyday life, we take it for granted that in interpersonal relationships, “I keep in mind your mind and you keep in mind mine” (p.375). This understanding of relationships as *a meeting of mental states* allows humans to experience a sense of personal connection when the other is physically absent or when the other shows intentions and goals of action that differ from ours, in complementary or even contrasting ways. Further, this background of an interpersonal connection allows us to appreciate the continuity of our current experience of sharing against previous encounters with the same person, characterized by different affective tones and motivations. This understanding is what allows *contingent* intersubjective exchanges to become an experience of relatedness. The complex level of intersubjectivity—indicating that “I know that you know that I know”—conveys a sense of mutuality and personal recognition and can only be reached gradually during ontogenesis, typically distinguishing the human species (Tomasello, 2019). As reported, earlier forms of intentional attributions that are based on simulation are insufficient to justify this complexity, and other mental prerequisites must be achieved by the infant. Indeed, current research shows that the quality of contingent intersubjective exchanges does not predict the quality of later intentional attributions (Moll et al., 2021): the new level of mentalistic understanding should be supported by a wider capacity—i.e., TOM (Baron-Cohen, 1991) or reflective functioning (Fonagy et al., 2002).

Developmental research has shown that early intentional attributions are present in early infancy but that they are better described as a teleological stance in which the understanding of intentions is directly derived from the outcome of the agent’s behavior (Gergely and Csibra, 2003). To overcome this

teleological thinking and to “decouple” internal states from outer reality, humans must become aware that actual behaviors are the consequence of an internal disposition that is conceived as intentions, thoughts, emotions, and desires. Developmental researchers have highlighted that such “intentional stance” or “second level intentionality” (Dennett, 1991) can only be reached through an objective representation or “second-order” representations of one’s own and others’ internal states (Fonagy et al., 2002). The first behavioral evidence of this developmental achievement is provided by the acquisition of an infant’s engagement in triadic (infant-other-object) joint attention interactions (Tomasello, 1995).

In human adults, joint attention skills are sustained by the dorsal region of the medial prefrontal cortex (mPFC) (Frith and Frith, 2006). Behavioral research has largely documented that infants are already able to discriminate between dyadic and triadic joint attention intercourses at age 3 months (Striano et al., 2005). Nevertheless, a full understanding of joint attention does not exist until age 9 months (Tomasello et al., 2005), perhaps by virtue of increased metabolic activity, which occurs in the frontal areas at approximately age 8 months (Chugani, 1994; Chugani et al., 1987).

EEG studies on the negative component (Nc) of event-related potentials (ERPs) support this assumption. The Nc is a negative deflection in frontocentral electrodes that is believed to reflect attentional orientation to salient stimuli (Richards et al., 2010; Striano et al., 2006) and attentional arousal (Soto-Icaza et al., 2015), because its amplitude is larger during sustained attention (Richards, 2003). In this domain, 9-month-old infants who engage in a joint attention interaction show higher amplitudes in the Nc of ERPs compared with non-joint attention intercourse (Striano et al., 2006).

According to the embodied simulation framework, ToM relies on the capacity to adopt a simulation routine that is, in turn, allowed by MNs (Gallese and Goldman, 1998). Nevertheless, as Frith and Frith (1999, 2001) have emphasized, the conscious reflection of one’s own mental states and those of others requires resources beyond the capacity to simulate or imitate an action and that are associated with the development of executive functions, especially inhibitory control (Carlson and Moses, 2001; Decety and Jackson, 2004).

Notably, the ToM reliably engages a network of brain regions that overlap partially with those that are involved in executive functions, including the mPFC, inferior frontal gyrus (IFG), IPL, and TPJ (Frith and Frith, 2012; Frolli et al., 2019; Gallagher and Frith, 2003; Koster-Hale and Saxe, 2013; Wade et al., 2018). Using near-infrared spectroscopy (fNIRS), Hyde et al. (2018) demonstrated that 7-month-old infants who stare at events evoking the ToM activate the TPJ but no other temporal regions (such as the superior temporal sulcus – STS) or the prefrontal regions (such as the mPFC). These data shed light on early organization of cerebral networks involved in the ToM, during infancy. At this early developmental stage, while the TPJ

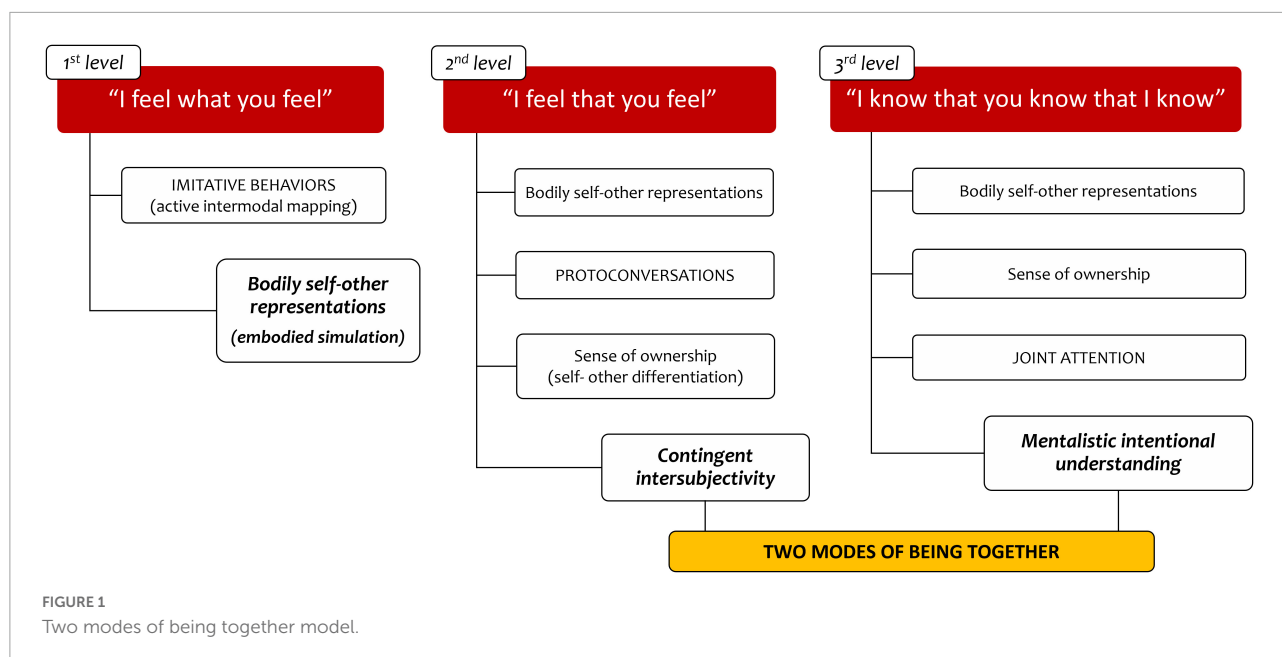
is already functionally organized for processing social stimuli that are relevant to the ToM, the mPFC (involved in inhibitory control; Draperi et al., 2022) might not be, due to its slow maturation during the first year of life (Chugani, 1994; Chugani et al., 1987).

Summary

Current neuroscience research is detailing early abilities that support the development of intersubjectivity. A putative and provisional reconstruction of what the human experiences in early social interactions can be obtained, relying on a multidisciplinary perspective (see Figure 1).

It appears that, during the first year of life, an infant can have an experience of sharing with others, which we define as “contingent intersubjectivity.” This form of connectedness (which can also be postulated in other species) is supported by neural mirror mechanisms and self-monitoring processes that enable the infant to distinguish internal from external sources of experience and to develop a sense of self based upon ownership and agency (at least on the level of interoceptive, tactile and sensory-motor information). We contend that this level of sharing possesses the prerequisites of intentional understanding and can thus be considered as a primary form of intersubjectivity. However, this form of intersubjective sharing seems to have some peculiarities that distinguish it from more mature forms of intersubjectivity.

In the first place, contingent intersubjectivity is temporarily limited to the ongoing interactions. At this level, the quality of relational experience is totally shaped by the actual affective and communicative exchanges between the infant and her caregivers, and no integration of such singular experiences can be achieved to establish the stable sense of connectedness that characterize interpersonal relationships. Secondly, contingent forms of sharing can be achieved only through actual correspondences between the infant and the caregiver’s intentions. As reported in psychodynamic literature (Weinberg and Tronick, 1997), when no such intentional attunement is reached, the sense of affective connection tends to decline. Furthermore, contingent intersubjectivity is self-referential, in that the infant, by default, feels the quality of the affective experience that is shared with the other as being hers. This means that, when the other exhibits disruptive or intrusive communicative behaviors, the infant tends to attribute to herself the caregiver’s negative internal states (Fotopoulou and Tsakiris, 2017; Lyons-Ruth et al., 2006). Thirdly, and most importantly, the experiential contents of contingent intersubjectivity are not formulated in terms of mentalistic explicit contents such as beliefs, thoughts, desires, emotions, intentions or goals. More peculiarly, the experience of contingent sharing is codified as a unitary form of experience, modeled by different sources of bodily and emotional information. Finally, it has to be



specified that the lack of the capacity for mentalization of such experiences of self-other interactions differentiates the basic form of intersubjective sharing from proper bi-personal connections, which are instead characteristic of more mature forms of human relatedness.

Implications for the dialog between psychoanalysis and neuroscience

We would like to stress that the reasoning we propose in this paper is also aimed at confronting two apparently incompatible ways of interpreting intersubjectivity, which is, the embodied simulation perspective and the mentalistic perspective. Specifically, our discussion highlighted that all experiences of intersubjective sharing are anchored to a bodily representation of interactions that comes into play both as a marker of self-other distinction and as a metaphorical bridge to enter and model the other's intentions and experience. As discussed above, the acquisition of the ability for mind-reading affords the individual a new and more ample way of experiencing interactions that is based on the capacity to live the current interactive exchanges as only one of the possible experiences of the relationship. We believe that these two tenets of our discussion bear some important theoretical and research consequences for the dialog between psychoanalysis and neuroscience.

One important aspect that needs to be clarified in future studies is the influence of early experiences of intersubjective sharing on the development of the mentalized forms of

relatedness. Research shows that impairments in the bodily self-monitoring may undermine reflective self-other representations (Tsakiris, 2017). At the same time, no clear continuity can be established between the quality of early intersubjective experience and future mentalizing capacities (Moll et al., 2021).

Far from falling back upon a reductionist perspective, current neuroscience research has identified the multifaceted nature of mental life, providing a new and articulated perspective on the relationships between the multifaceted aspects of the individual experience of the self in relation to her social environment. The contributions included in this paper clearly evidence the bodily foundation of human psychic life, as affirmed by the psychoanalytic thinking from its very beginning (Freud, 1914; Heimann et al., 1952). At the same time, psychoanalysis has always dealt with the necessity to understand the psychic principles that lead the transformation of the unaware experience deriving from bodily functioning into conscious thoughts and emotional contents (Bion, 1962). Therefore, for both researchers and psychoanalysts, an important area of investigation definitively concerns the kind of mutual influence between the representational codes and the quality of experience pertaining to the contingent and reflective modes of intersubjective experiences. Investigations into the neural networks supporting each aspect of emotional sharing indeed point to the issue of how the brain functioning reaches the integration between bottom-up and top-down processing of self and relationships, leading to personal meaning (Northoff and Scalabrini, 2021).

We believe that a new important step for psychoanalytic investigations—as led by neuro-psychoanalysis and neuroscience—has already been achieved. The affirmation

of bi-personal models in current psychoanalytic thinking (as opposed to the mono-personal framework of classic psychoanalysis) has correctly highlighted the importance of real social experiences for the development of personality and psychopathology (Gill, 1994). Current psychoanalytic relational orientation as well as infant research and attachment theory (Bowlby, 1969/1982) posit that the human fundamental motivation relies in the seek to establish a harmonic relationship with the caregivers, who are to meet this relational trust by offering recognition, regulation, sharing, and protection to the infant's biological and emotional needs (Eagle, 2011). Coherently, it is also postulated that the positive experience of early relationships *de facto* entails a natural equivalence between the individual's needs, her affective experience, and the object's presence. It seems that such a polarized view runs the risk of throwing away the psychoanalytic baby with the bathwater.

Current neuroscience research seems to afford a more complex view of human relatedness. The presumed infant's capacity to realistically appreciate and experience interpersonal relationships as such seems indeed the result of a progressive construction, depending on both maturation and experience. Early forms of relatedness are bound to a bodily experience constituting the idiosyncratic and personal meaning that shapes individual sense of the self in relation to the other (Modell, 1993). It should not be underestimated that psychoanalytic observations have deserved recognitions for having connected psychopathological states to the effects produced by the incompatibility between the underworld of private representations and the more mature and articulated experiences of relationships.

Possible implications for psychoanalytic theory and clinical models

In this regard, we contend that the multi-layered account of the development of intersubjective capacities might help modeling some of the clinical phenomena envisioned by classic psychoanalysis, in an updated theoretical framework.

By virtue of the acquisition of mentalistic abilities, the creation of a virtual internal space that belongs to the self and the other allows the individual to experience that she and the other have different feelings and thoughts about the world, and that they can come to share a unique experience or re-establish a connection when this connection is lost. These achievements allow the individual to stay within and without the current experience of the self and the other. Notably, neuroscience research also shows that this developmental conquest does not rule out the embodied knowledge mechanisms that allow for

primary forms of sharing and perceptions of another individual's intentions. Neural mirror mechanisms, along with self-monitoring of bodily activation occurring in the interaction, constitute the basis for self-other experience, representing a "metaphorical bridge" for interpreting the meaning of ongoing relationships throughout life (Modell, 2003).

We suggest that our theoretical proposal might also shed a new light on some classic psychoanalytic observations of clinical phenomena observed, for instance, in borderline conditions. The "lack of constancy of the libidinal objects" (Mahler et al., 1975) and the ensuing fears of abandonment and symbiotic swallowing characterizing these patients may be read as the difficulty to integrate the contingent intersubjective experience into a mentalistic understanding of relationships. On the one hand, the fall in the mentalizing abilities may lead these patients to experience the absence of any contingent response on the part of the other as a definitive loss of the relationship, generating a state of extreme and hopeless solitude. Moreover, the ensuing strong thrust to regain a sense of relationship may lead these patients to evoke forms of contingent exchanges with the other, in which past or present experiences of rejection, frustration or intrusion may take over. The self-referential and all-encompassing mode accompanying contingent exchanges may plunge these patients in a state of self-other confusion, in which the sense of personal identity is lost. In a similar vein, it has been suggested that the feeling of entrapment into the repetition of a "traumatic script" (Meares, 2012), the identification with the traumatic object, and the difficulty to disentangle the self-experience from that of the object showed by borderline patients (Gabbard et al., 2006) may be related to the activation of memories of contingent interactions with the traumatizing figures, that are not sustained by an adequate cortical activity of second-order representations (Fonagy et al., 2002).

We would finally like to stress how the multi-layered view of intersubjective experiences emerging from the neuroscience and neuropsychanalytic investigations bears some important consequences for the conception of treatment. Some recent psychoanalytic orientations (Bateman and Fonagy, 2013) have emphasized the importance to help the patient to mentalize her dispositional states to achieve a good therapeutic outcome. Undoubtedly, enhancing the patient's reflective functioning is the unavoidable tool of work for any psychodynamically inspired psychotherapy. Nevertheless, the perspective presented in this contribution also points to the necessity to bring reflection into the field of the embodied experience of relationship. The creation of a new awareness of oneself should not be thought of as an epistemological effort (Ogden, 2019). Rather, reflection and interpretation should directly address those idiosyncratic bodily experiences of the relationship that constitute the building block and raw material upon which the

sense of being a person in relation to the world comes to life and must be constantly re-created.

Conclusion

In this paper, it has been proposed that neuroscience and psychoanalysis now converge in showing that there are at least two modes of being together. The multi-colored emotions and perceptions we live from within our body in the encounter with the others is what tells us how it feels to be in a relationship and to live a life. At the same time, we must or should be able to see and think about ourselves from without, to reach and maintain the sense of being a person, no matter what the ongoing interaction is. The contingent mode of being together provides the emotions and sensations that render the relationships meaningful and worth-living, while the reflective mode of being together allows us to create a history of these meaningful relationship, for good and for bad. There is always price that we pay to live relationships from both perspectives. It is the possibility to oscillate and compound these two modes of being together that makes us humans.

References

- Atzil, S., Gao, W., Fradkin, I., and Barrett, L. F. (2018). Growing a social brain. *Nat. Hum. Behav.* 2, 624–636. doi: 10.1038/s41562-018-0384-6
- Atzil, S., Hendler, T., and Feldman, R. (2011). Specifying the neurobiological basis of human attachment: 690 brain, hormones, and behavior in synchronous and intrusive mothers. *Neuropsychopharmacology* 36, 2603–2615. doi: 10.1038/npp.2011.172
- Avenanti, A., Buetti, D., Galati, G., and Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nat. Neurosci.* 8, 955–960. doi: 10.1038/nn1481
- Baron-Cohen, S. (1991). “Precursors to a theory of mind: Understanding attention in others,” in *Natural theories of mind: Evolution, development and simulation of everyday mindreading*, ed. A. Whiten (Oxford: Basil Blackwell), 233–251.
- Bateman, A., and Fonagy, P. (2013). Mentalization-based treatment. *Psychoanal. Inq.* 33, 595–613. doi: 10.1080/07351690.2013.835170
- Beebe, B., and Lachmann, F. M. (2002). *Infant research and adult treatment: A dyadic systems approach*. New York, NY: The Analytic Press.
- Bekoff, M. (2007). *The emotional lives of animals: A leading scientist explores animal joy, sorrow, and empathy—and why they matter*. Novato, CA: New World Library.
- Bion, W. (1962). *Learning from experience*. London: William Heinman Medical Books.
- Bowlby, J. (1969/1982). *Attachment and loss: Vol. 1. Attachment*, 2nd Edn. New York, NY: Basic Books.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by non-conspecifics: An fMRI study. *J. Cogn. Neurosci.* 16, 114–126. doi: 10.1162/089892904322755601
- Cabanis, M., Pyka, M., Mehl, S., Müller, B. W., Loos-Jankowiak, S., Winterer, G., et al. (2013). The precuneus and the insula in self-attributional processes. *Cogn. Affect. Behav. Neurosci.* 13, 330–345. doi: 10.3758/s13415-012-0143-5
- Caramazza, A., Anzellotti, S., Strnad, L., and Lingnau, A. (2014). Embodied cognition and mirror neurons: A critical assessment. *Annu. Rev. Neurosci.* 14, 1–15. doi: 10.1146/annurev-neuro-071013-013950
- Carey, S., and Spelke, E. (1996). Science and core knowledge. *Philos. Sci.* 63, 515–533. doi: 10.1086/289971
- Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children’s theory of mind. *Child Dev.* 72, 1032–1053. doi: 10.1111/1467-8624.00333
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502. doi: 10.1073/pnas.0935845100
- Carruthers, P., and Smith, P. K. (1996). *Theories of theory of mind*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511597985
- Cermolacce, M., Naudin, J., and Parnas, J. (2007). The “minimal self” in psychopathology: Re-examining the self-disorders in the schizophrenia spectrum. *Conscious. Cogn.* 16, 703–714. doi: 10.1016/j.concog.2007.05.013
- Chan, Y. C., and Lavalley, J. P. (2015). Temporo-parietal and fronto-parietal lobe contributions to theory of mind and executive control: An fMRI study of verbal jokes. *Front. Psychol.* 6:1285. doi: 10.3389/fpsyg.2015.01285
- Chugani, H. T. (1994). “Development of regional brain glucose metabolism in relation to behavior and plasticity,” in *Human behavior and the developing brain*, eds G. Dawson and K. W. Fischer (New York, NY: The Guilford Press), 153–175.
- Chugani, H. T., Phelps, M. E., and Mazziotta, J. C. (1987). Positron emission tomography study of human brain functional development. *Ann. Neurol.* 22, 487–497. doi: 10.1002/ana.410220408
- Cuevas, K., Cannon, E. N., Yoo, K., and Fox, N. A. (2014). The infant EEG mu rhythm: Methodological considerations and best practices. *Dev. Rev.* 34, 26–43. doi: 10.1016/j.dr.2013.12.001
- Damasio, A. R. (2010). *Self comes to mind: Constructing the conscious brain*. New York, NY: Random House.

Author contributions

RW conceived the work. Both authors searched and studied the literature, wrote the article, and contributed to the viewpoint that the article expresses.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Davey, C. G., Pujol, J., and Harrison, B. J. (2016). Mapping the self in the brain's default mode network. *Neuroimage* 132, 390–397. doi: 10.1016/j.neuroimage.2016.02.022
- de Waal, F. W. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674033177
- Decety, J., and Chaminade, T. (2003). When self represents the other: A new cognitive neuroscience view on psychological identification. *Conscious. Cogn.* 12, 577–596. doi: 10.1016/S1053-8100(03)00076-X
- Decety, J., and Ickes, W. (2009). *The social neuroscience of empathy*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/9780262012973.001.0001
- Decety, J., and Jackson, P. L. (2004). The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71–100. doi: 10.1177/1534582304267187
- Dennett, D. (1991). *Consciousness explained*. Boston, MA: Little Brown.
- Di Bernardo, M. (2021). Neuropsychology and intersubjectivity: An interdisciplinary approach. *Axiomathes* 1–17. doi: 10.1007/s10516-021-09590-3
- Dixon, M. L., Moodie, C. A., Goldin, P. R., Farb, N., Heimberg, R. G., Zhang, J., et al. (2022). Frontoparietal and default mode network contributions to self-referential processing in social anxiety disorder. *Cogn. Affect. Behav. Neurosci.* 22, 187–198. doi: 10.3758/s13415-021-00933-6
- Draperi, M., Aïte, A., Cassotti, M., Le Stanc, L., Houdé, O., and Borst, G. (2022). Development of cool and hot theory of mind and cool and hot inhibitory control abilities from 3.5 to 6.5 years of age. *PLoS One* 17:e0262251. doi: 10.1371/journal.pone.0262251
- Eagle, M. (2011). *From classic to contemporary psychoanalysis: An integration and a critique*. London: Routledge. doi: 10.4324/9780203868553
- Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., and Jeannerod, M. (2003). Modulating the experience of agency. *Neuroimage* 18, 324–333. doi: 10.1016/S1053-8119(02)00041-1
- Fonagy, P., Gergely, G., Target, M., and Jurist, E. (2002). *Mentalization, affective regulation and the development of the self*. New York, NY: Other Press.
- Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychanalysis* 19:775. doi: 10.1080/15294145.2017.1294031
- Freud, S. (1914). “On narcissism: An introduction,” in *The standard edition of the complete psychological works of Sigmund Freud*, Vol. XIV, ed. J. Strachey (London: The Hogarth Press).
- Frith, C. D., and Frith, U. (1999). Interacting minds—a biological basis. *Science* 286, 1692–1695. doi: 10.1126/science.286.5445.1692
- Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron* 50, 531–534. doi: 10.1016/j.neuron.2006.05.001
- Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449
- Frith, U., and Frith, C. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155. doi: 10.1111/1467-8721.00137
- Frolli, A., La Penna, I., Cavallaro, A., and Ricci, M. C. (2019). Theory of mind: Autism and typical development. *Acad. J. Ped. Neonatol.* 8:555799. doi: 10.19080/AJPN.2019.08.555799
- Gabbard, G. O., Miller, L. A., and Martinez, M. (2006). “A neurobiological perspective on mentalizing and internal object relations in traumatized patients with borderline personality disorder,” in *Handbook of mentalization-based treatment*, eds J. G. Allen and P. Fonagy (Chichester: Wiley), 123–140. doi: 10.1002/9780470712986.ch5
- Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of ‘theory of mind’. *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5
- Gallagher, S. (2005). “Dynamic models of body schematic processes,” in *Body image and body schema*, eds H. De Preester and V. Knockaert (Amsterdam: John Benjamins Publishers), 233–250. doi: 10.1075/aicr.62.15gal
- Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 171–180. doi: 10.1159/000072786
- Gallese, V. (2006). Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain Res.* 1079, 15–24. doi: 10.1016/j.brainres.2006.01.054
- Gallese, V. (2014). “From mirror neurons to the shared manifold hypothesis: A neurophysiological account of intersubjectivity,” in *Biology and knowledge revisited: From neurogenesis to psychogenesis*, eds S. T. Parker, J. Langer, and C. Milbrath (London: Routledge), 197–222.
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* 119, 593–609. doi: 10.1093/brain/119.2.593
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/S1364-6613(98)01262-5
- Gallup, G. G. (1998). Self-awareness and the evolution of social intelligence. *Behav. Process.* 42, 239–247. doi: 10.1016/S0376-6357(97)00079-X
- Gergely, G., and Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends Cogn. Sci.* 7, 287–292. doi: 10.1016/S1364-6613(03)00128-1
- Gill, M. M. (1994). *Psychoanalysis in transition: A personal view*. Hillsdale, NJ: The analytic Press.
- Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mind reading*. New York, NY: Oxford University Press. doi: 10.1093/0195138929.001.0001
- Grossmann, T., Oberecker, R., Koch, S. P., and Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858. doi: 10.1016/j.neuron.2010.03.001
- Heimann, P., Isaacs, S., Klein, M., and Riviere, J. (1952). *Developments in psychoanalysis*. London: The Hogarth Press.
- Hakuno, Y., Hata, M., Naoi, N., Hoshino, E. I., and Minagawa, Y. (2020). Interactive live fNIRS reveals engagement of the temporoparietal junction in response to social contingency in infants. *Neuroimage* 218:116901. doi: 10.1016/j.neuroimage.2020.116901
- Heyes, C., and Catmur, C. (2022). What happened to mirror neurons? *Perspect. Psychol. Sci.* 17, 153–168. doi: 10.1177/1745691621990638
- Hyde, D. C., Simon, C. E., Ting, F., and Nikolaeva, J. I. (2018). Functional organization of the temporal-parietal junction for theory of mind in preverbal infants: A near-infrared spectroscopy study. *J. Neurosci.* 38, 4264–4274. doi: 10.1523/JNEUROSCI.0264-17.2018
- Jacobson, E. (1964). *The self and the object world*. New York, NY: International Universities Press, Inc.
- Jamieson, D., and Bekoff, M. (1996). *Readings in animal cognition*. Cambridge, MA: The MIT Press.
- Kernberg, O. (1985). *Borderline conditions and pathological narcissism*. New York, NY: Jason Aronson Books.
- Klein, M. (1946). Notes on some schizoid mechanisms. *Int. J. Psychoanal.* 27, 99–110. doi: 10.4324/9780429473661-9
- Kohut, H. (1971). *The analysis of the self: A systematic approach to the psychoanalytic treatment of narcissistic personality disorders*. Chicago, IL: University of Chicago Press.
- Koster-Hale, J., and Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron* 79, 836–848. doi: 10.1016/j.neuron.2013.08.020
- Lyons-Ruth, K., Dutra, L., Schuder, M. R., and Bianchi, I. (2006). From infant attachment disorganization to adult dissociation: Relational adaptations or traumatic experiences? *Psychiatr. Clin. North Am.* 29, 63–86. doi: 10.1016/j.psc.2005.10.011
- Mahler, M., Pine, M., and Bergman, A. (1975). *The psychological Birth of the Human Infant*. New York, NY: Basic Books.
- Marraffa, M., and Meini, C. (2019). A social-cognitive approach to self-construction: From Bowlby to Piaget and back again. *Paradigmi* 37, 121–140.
- Marshall, P. J., Bar-Haim, Y., and Fox, N. A. (2002). Development of the EEG from 5 months to 4 years of age. *J. Clin. Neurophysiol.* 113, 1199–1208. doi: 10.1016/S1388-2457(02)00163-3
- Marshall, P. J., and Meltzoff, A. N. (2011). Neural mirroring systems: Exploring the EEG mu rhythm in human infancy. *Dev. Cogn. Neurosci.* 1, 110–123. doi: 10.1016/j.dcn.2010.09.001
- Marshall, P. J., and Meltzoff, A. N. (2014). Neural mirroring mechanisms and imitation in human infants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130620. doi: 10.1098/rstb.2013.0620
- Meltzoff, A. N., and Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behav. Dev.* 15, 479–505. doi: 10.1016/0163-6383(92)80015-M
- Marshall, P. J., Saby, J. N., and Meltzoff, A. N. (2013). Imitation and the developing social brain: Infants somatotopic EEG patterns for acts of self and other. *Int. J. Psychol. Res.* 6, 22–29. doi: 10.21500/20112084.714
- Meares, R. (2012). *The dissociation model of borderline personality disorder*. New York, NY: Norton Professional Books.

- Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Dev. Psychol.* 24, 470–476. doi: 10.1037/0012-1649.24.4.470
- Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Dev. Sci.* 10, 126–134. doi: 10.1111/j.1467-7687.2007.00574.x
- Meltzoff, A. N. (2013). "Origins of social cognition: Bidirectional self-other mapping and the 'Like-Me' hypothesis," in *Navigating the social world: What infants, children, and other species can teach us*, eds M. R. Banaji and S. A. Gelman (New York, NY: Oxford University Press), 139–144. doi: 10.1093/acprof:oso/9780199890712.003.0025
- Meltzoff, A. N., and Marshall, P. J. (2018). Human infant imitation as a social survival circuit. *Curr. Opin. Behav. Sci.* 24, 130–136. doi: 10.1016/j.cobeha.2018.09.006
- Meltzoff, A. N., and Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behav. Dev.* 17, 83–99. doi: 10.1016/0163-6383(94)90024-8
- Meltzoff, A. N., and Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Infant Child Dev.* 6, 179–192. doi: 10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R
- Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends Cogn. Sci.* 15, 483–506.
- Modell, A. H. (1993). *The private self*. Cambridge, MA: Harvard University Press.
- Modell, A. H. (2003). *Imagination and the meaningful brain*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/3666.001.0001
- Moll, H., Püschel, E., Ni, Q., and Little, A. (2021). Sharing experiences in infancy: From primary intersubjectivity to shared intentionality. *Front. Psychol.* 12:667679. doi: 10.3389/fpsyg.2021.667679
- Nagy, E., and Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behav. Dev.* 27, 54–63. doi: 10.1016/j.infbeh.2003.06.004
- Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., and Panksepp, J. (2006). Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *Neuroimage* 31, 440–457. doi: 10.1016/j.neuroimage.2005.12.002
- Northoff, G., and Scalabrini, A. (2021). Project for a spatiotemporal neuroscience. Brain and psyche share their topography and dynamic. *Front. Psychol.* 12:717402. doi: 10.3389/fpsyg.2021.717402
- Ogden, T. H. (2019). Ontological psychoanalysis or "What do you want to be when you grow up?" *Psychoanal. Q.* 88, 661–684. doi: 10.1080/00332828.2019.1656928
- Park, H.-D., and Blanke, O. (2019). Coupling inner and outer body for self-consciousness. *Trends Cogn. Sci.* 23, 377–388. doi: 10.1016/j.tics.2019.02.002
- Peterfreund, E. (1978). Some critical comments on the psychoanalytic conceptualizations of infancy. *Int. J. Psychoanal.* 59, 427–441.
- Qin, P., Wang, M., and Northoff, G. (2020). Linking bodily, environmental and mental states in the self – a three-level model based on a meta-analysis. *Neurosci. Biobehav. Rev.* 115, 77–95. doi: 10.1016/j.neubiorev.2020.05.004
- Richards, J. E. (2003). Attention affects the recognition of briefly presented visual stimuli in infants: An ERP study. *Dev. Sci.* 6, 312–328. doi: 10.1111/1467-7687.00287
- Richards, J. E., Reynolds, G. D., and Courage, M. L. (2010). The neural bases of infant attention. *Curr. Dir. Psychol. Sci.* 19, 41–46. doi: 10.1177/0963721409360003
- Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141. doi: 10.1016/0926-6410(95)00038-0
- Rizzolatti, G., and Fogassi, L. (2014). The mirror mechanism: Recent findings and perspectives. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130420. doi: 10.1098/rstb.2013.0420
- Rizzolatti, G., and Sinigaglia, C. (2007). Mirror neurons and motor intentionality. *Funct. Neurol.* 22, 205–210.
- Rizzolatti, G., and Sinigaglia, C. (2008). Further reflections on how we interpret the actions of others. *Nature* 455:589. doi: 10.1038/455589b
- Rochat, P., and Striano, T. (2000). Perceived self in infancy. *Infant Behav. Dev.* 23, 513–530. doi: 10.1016/S0163-6383(01)00055-8
- Saby, J. N., Meltzoff, A. N., and Marshall, P. J. (2013). Infants' somatotopic neural responses to seeing human actions: I've got you under my skin. *PLoS One* 8:e77905. doi: 10.1371/journal.pone.0077905
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040. doi: 10.1073/pnas.0805234105
- Schore, A. N. (2021). The interpersonal neurobiology of intersubjectivity. *Front. Psychol.* 12:648616. doi: 10.3389/fpsyg.2021.648616
- Schurz, M., Aichhorn, M., Martin, A., and Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: A meta-analysis of functional brain imaging studies. *Front. Hum. Neurosci.* 7:712. doi: 10.3389/fnhum.2013.00712
- Schurz, M., Tholen, M. G., Perner, J., Mars, R. B., and Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities. *Hum. Brain Mapp.* 38, 4788–4805. doi: 10.1002/hbm.23675
- Simpson, E. A., Murray, L., Paukner, J., and Ferrari, P. F. (2014). The mirror neuron system as revealed through neonatal imitation: Presence from birth, predictive power and evidence of plasticity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130289. doi: 10.1098/rstb.2013.0289
- Singer, T., and Frith, C. (2005). The painful side of empathy. *Nat. Neurosci.* 8, 845–846. doi: 10.1038/nn0705-845
- Soto-Icaza, P., Aboitiz, F., and Billeke, P. (2015). Development of social skills in children: Neural and behavioral evidence for the elaboration of cognitive models. *Front. Neurosci.* 9:333. doi: 10.3389/fnins.2015.00333
- Southgate, V., Johnson, M. H., Karoui, I. E., and Csibra, G. (2010). Motor system activation reveals infants' on-line prediction of others' goals. *Psychol. Sci.* 21, 355–359. doi: 10.1177/0956797610362058
- Southgate, V., Johnson, M. H., Osborne, T., and Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biol. Lett.* 5, 769–772. doi: 10.1098/rsbl.2009.0474
- Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Dev. Sci.* 10, 89–96. doi: 10.1111/j.1467-7687.2007.00569.x
- Sterling, P., and Laughlin, S. (2015). *Principles of neural design*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/9780262028707.001.0001
- Stern, D. N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York, NY: Basic Books.
- Stern, D. N. (2004). *The present moment in psychotherapy and everyday life*. New York, NY: Basic Books.
- Stern, D. N. (2005). "Intersubjectivity," in *Textbook of psychoanalysis*, eds E. S. Person, A. M. Cooper, and G. O. Gabbard (Arlington, VA: American Psychiatric Publishing, Inc), 77–92.
- Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiol. Behav.* 106, 5–15. doi: 10.1016/j.physbeh.2011.06.004
- Striano, T., Henning, A., and Stahl, D. (2005). Sensitivity to social contingencies between 1 and 3 months of age. *Dev. Sci.* 8, 509–518. doi: 10.1111/j.1467-7687.2005.00442.x
- Striano, T., Reid, V. M., and Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *Eur. J. Neurosci.* 23, 2819–2823. doi: 10.1111/j.1460-9568.2006.04822.x
- Suzuki, K., Garfinkel, S. N., Critchley, H. D., and Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia* 51, 2909–2917. doi: 10.1016/j.neuropsychologia.2013.08.014
- Tomasello, M. (1995). "Joint attention as social cognition," in *Joint attention: Its origins and role in development*, eds C. Moore and P. J. Dunham (New York, NY: Psychology Press), 103–130.
- Tomasello, M. (1999). *The cultural origins of human intelligence*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2019). *Becoming humans: A theory of ontogeny*. Cambridge, MA: Harvard University Press. doi: 10.4159/9780674988651
- Tomasello, M., Call, J., and Hare, B. (2003). Chimpanzees understand psychological states—the question is which ones and to what extent. *Trends Cogn. Sci.* 7, 153–156. doi: 10.1016/S1364-6613(03)00035-4
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.* 28, 675–735. doi: 10.1017/S0140525X05000129
- Tomasello, M., and Call, J. (1997). *Primate cognition*. Oxford: Oxford University Press.
- Trevarthen, C. (1974). The psychology of speech development. *Neurosci. Res. Prog. Bull.* 12, 570–585. doi: 10.1152/ajpregu.1984.246.6.R888
- Trevarthen, C. (1979). "Communication and cooperation in early infancy: A description of primary intersubjectivity," in *Before speech: The beginning of*

interpersonal communication, ed. M. Bullock (Cambridge: Cambridge University Press), 321–347.

Trevarthen, C. (1993). “The self born in intersubjectivity: The psychology of an infant communicating,” in *The perceived self: Ecological and interpersonal sources of self-knowledge*, ed. U. Neisser (New York, NY: Cambridge University Press), 121–173. doi: 10.1017/CBO9780511664007.009

Trevarthen, C. (1998). “The concept and foundations of infant intersubjectivity,” in *Intersubjective communication and emotion in early ontogeny*, ed. S. Braten (New York, NY: Cambridge University Press), 15–46.

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Q. J. Exp. Psychol.* 70, 597–609. doi: 10.1080/17470218.2016.1181768

Wade, M., Prime, H., Jenkins, J. M., Yeates, K. O., Williams, T., and Lee, K. (2018). On the relation between theory of mind and executive functioning: A developmental cognitive neuroscience perspective. *Psychon. Bull. Rev.* 25, 2119–2140. doi: 10.3758/s13423-018-1459-0

Weinberg, M. K., and Tronick, E. Z. (1997). “Depressed mothers and infants: Failure to form dyadic states of consciousness,” in *Postpartum depression and child development*, eds L. Murray and P. J. Cooper (New York, NY: Guilford Press), 54–81.

Winnicott, D. W. (1945). Primitive emotional development. *Int. J. Psychoanal.* 26, 137–143. doi: 10.1111/1745-8315.12083



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Sara Masland,
Pomona College, United States
Ondrej Pec,
Charles University, Czechia

*CORRESPONDENCE

Juan Pablo Jiménez
jjimenez@med.uchile.cl

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 27 May 2022

ACCEPTED 18 August 2022

PUBLISHED 12 September 2022

CITATION

Quevedo Y, Booij L, Herrera L,
Hernández C and Jiménez JP
(2022) Potential epigenetic
mechanisms in psychotherapy: a pilot
study on DNA methylation and
mentalization change in borderline
personality disorder.
Front. Hum. Neurosci. 16:955005.
doi: 10.3389/fnhum.2022.955005

COPYRIGHT

© 2022 Quevedo, Booij, Herrera,
Hernández and Jiménez. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Potential epigenetic mechanisms in psychotherapy: a pilot study on DNA methylation and mentalization change in borderline personality disorder

Yamil Quevedo^{1,2}, Linda Booij^{3,4}, Luisa Herrera⁵,
Cristobal Hernández^{2,6} and Juan Pablo Jiménez^{1,2*}

¹Departamento de Psiquiatría y Salud Mental Oriente, Universidad de Chile, Santiago, Chile,

²Millenium Institute for Depression and Personality Research, Santiago, Chile, ³Department of Psychology, Concordia University, Montreal, QC, Canada, ⁴Sainte-Justine Hospital Research Centre, Montreal, QC, Canada, ⁵Programa de Genética Humana, Instituto de Ciencias Biomédicas, Universidad de Chile, Santiago, Chile, ⁶Escuela de Psicología, Universidad Adolfo Ibañez, Santiago, Chile

Genetic and early environmental factors are interwoven in the etiology of Borderline Personality Disorder (BPD). Epigenetic mechanisms offer the molecular machinery to adapt to environmental conditions. There are gaps in the knowledge about how epigenetic mechanisms are involved in the effects of early affective environment, development of BPD, and psychotherapy response. We reviewed the available evidence of the effects of psychotherapy on changes in DNA methylation and conducted a pilot study in a sample of 11 female adolescents diagnosed with BPD, exploring for changes in peripheral DNA methylation of *FKBP5* gene, which encodes for a stress response protein, in relation to psychotherapy, on symptomatology and underlying psychological processes. For this purpose, measures of early trauma, borderline and depressive symptoms, psychotherapy outcome, mentalization, and emotional regulation were studied. A reduction in the average *FKBP5* methylation levels was observed over time. Additionally, the decrease in *FKBP5* methylation observed occurred only in those individuals who had early trauma and responded to psychotherapy. The results suggest an effect of psychotherapy on epigenetic mechanisms associated with the stress response. The finding that epigenetic changes were only observed in patients with early trauma suggests a specific molecular mechanism of recovery. The results should be taken with caution given the small sample size. Also, further research is needed to adjust for confounding factors and include endocrinological markers and therapeutic process variables.

KEYWORDS

borderline personality disorder, DNA methylation, mentalization, psychotherapy, epigenetics

Introduction

Borderline Personality Disorder (BPD) is characterized by a general pattern of instability in affect regulation, impulse control, interpersonal relationships, and self-image. A revision of the data from The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) study in the United States estimates a lifetime prevalence of 2.7% (Trull et al., 2010). In mental health settings, the prevalence of BPD is expected to be 10% in outpatient and between 15% and 25% in inpatients (Leichsenring et al., 2011). The prevalence of BPD is relatively similar in adolescents and adults, and presents acute symptomatology, such as suicidal ideation, impulsive behaviors, and non-suicidal self-injury (NSSI; Stead et al., 2019). Adolescents with BPD have academic difficulties and social relationship problems (Kaess et al., 2017). Additionally, this disorder is a significant predictor of substance use and mood disorders (Chanen et al., 2008). Individuals with BPD symptoms at mean age of 14 years show lower life satisfaction, more general impairment, and need for services at mean age of 33 (Winograd et al., 2008).

Given the implications of BPD on the functioning and mental health of adolescents, a better understanding of the interplay between genetic and environmental factors that contribute to their symptomatic expression is highly needed.

The development of BPD pathology implies a complex etiopathogenic trajectory in which adverse early life events in the presence of genetic susceptibility that confer sensitivity to the environment, can lead to the development over time of the BPD phenotype or its underlying traits (Gunderson and Lyons-Ruth, 2008; Bulbena-Cabre et al., 2018). However, evidence of the underlying molecular mechanisms is largely unknown. A large-scale family study estimates the heritability of the disorder at 46% (95% CI = 39–53; Skoglund et al., 2021). Despite evidence from twin studies showing that BPD is highly heritable, genetic association studies are so far inconclusive (Calati et al., 2013; Amad et al., 2014).

Epigenetic processes are sensitive to environmental conditions and can operate as mechanisms that allow early environmental experiences to trigger phenotypic modifications without modifying the genotype (Weaver, 2007). A limited number of studies have shown relations between DNA methylation patterns and the presence of childhood stress in individuals with BPD symptoms, showing associations with genes involved in stress regulation and neuroplasticity. A positive correlation has been observed between the levels of DNA methylation of exon 1F the Glucocorticoid Receptor gene (NR3C1) promoting region, child maltreatment (physical abuse), and clinical severity in a sample of individuals with BPD (Martín-Blanco et al., 2014). Perroud et al. (2013) examined 115 individuals diagnosed with borderline personality disorder (BPD). All received intensive dialectical behavioral therapy.

Those who responded to treatment showed a decrease in the percentage of exon I and IV *BDNF* DNA methylation. Changes in the methylation status were significantly related to changes in depressive symptoms, hopelessness, and impulsivity.

To date, there is no psychopharmacological treatment with robust evidence of effectiveness for the treatment of BPD. On the other hand, various types of specialized psychotherapy have shown efficacy in reducing symptoms and improving global functioning (Choi-Kain et al., 2017). These effects could occur by modification of the epigenetic profile. In line with the preceding, there are few recent studies relating BPD and the potential epigenetic effect of psychotherapeutic treatments, specifically on *BDNF* gene DNA methylation (Perroud et al., 2013; Thomas et al., 2018) and *APBA3* and *MCF2* genes DNA methylation (Knoblich et al., 2018). However, the exploration of the association between the change produced at the level of symptoms or personality functioning induced by psychotherapeutic interventions and changes at the molecular level is still very scarce (Jiménez et al., 2018).

Accordingly, the aim of the present study is to discuss the role of epigenetic changes as a mechanism of gene-environment interaction, its relevance for BPD, and current evidence on DNA methylation changes in individuals with BPD during psychotherapy. Second, the concept of mentalization as a capacity for the processing of the interpersonal context, its development from the early experiences of care, and its participation as a possible common factor in psychotherapy in BPD is exposed. Finally, the notion that subjective processing of the social environment can act on the genetic expression as a mechanism of adaptation of the BPD phenotype is discussed.

Epigenetic processes as a mechanism of gene-environment interaction

Epigenetic modifications refer to stable alterations of potential gene expression during development and cell proliferation, that are held through cell divisions and do not alter the DNA sequence (Jaenisch and Bird, 2003). These correspond to heritable patterns of DNA methylation and hydroxymethylation, posttranslational histone modifications, and gene expression regulation by non-coding RNAs (Zannas et al., 2015). The combination of these changes determines a specific pattern of gene expression, which is highly dynamic and permeable to environmental influences. It is also heterogeneous in different organisms, tissues, and cell types, and changes according to the stages of development. Therefore, they correspond to a complex set of mechanisms of

“phenotypic plasticity” in response to environmental demands (Ecker et al., 2018).

Experimental models have studied the impact of early adversity as a function of maternal care. For instance, in rats, the effect of maternal care behavior like licking, grooming (LG), and back arch-nursing (ABN) on behavior of the offspring and DNA methylation in the glucocorticoid receptor gene (*NR3C1*) have been reported (Lutz and Turecki, 2014). Increased DNA methylation in promoter regions of GR gene (*NR3C1*; hence more inactive chromatin and therefore lower transcription) in the hippocampus of adult rats reared by mothers with low levels of LG-ABN compared to offspring reared by mothers with high LG-ABN was observed. This lower GR expression was associated with less negative feedback in the HPA axis and higher reactivity to stress (Lutz and Turecki, 2014).

In rats exposed to early stress (separation of mother and calf), increased secretion of corticosterone and a persistent increase in Arginine Vasopressin (AVP) expression in neurons of the paraventricular nucleus of the Hippocampus are observed. AVP acts by enhancing the action of Corticotropin Releasing Hormone (CRH) under sustained stress situations. This increase is associated with hypomethylation in the regulatory region CGI3 of *AVP* gene and with altered behaviors of stress coping (Murgatroyd et al., 2009).

In humans, a number of studies have explored the relationship between early adverse environment and changes in methylation patterns, using candidate genes and epigenome-wide strategies.

Regarding studies linking candidate genes studies related to the stress response and adverse events in childhood, a significant correlation between the number of adverse events reported and one exon 1F DNA methylation site (cg17860381, located in exon 1F) of *NR3C1* gene was observed in lymphocytes of females who reported childhood adverse events, including physical, emotional and sexual abuse. Moreover, this pattern of adverse events and DNA methylation was correlated with borderline symptoms (Radtke et al., 2015).

A systematic review conducted by Turecki and Meaney (2016) regarding the effects of social environment on *NR3C1* gene methylation in humans showed that there was a consistent relationship (16 out of 17 reviewed studies) between early life adversity and increased exon 1F DNA methylation across different tissues (blood, saliva, buccal cells, and brain tissue). However, the results are inconsistent when exploring the association between exon 1F methylation and psychopathology including post-traumatic stress disorder and Depression. These results are also inconsistent when exploring methylation in other sites of exon 1 of the same gene in relation to early adversity, suggesting the need for further research to determine the permeability and stability of DNA methylation of each specific site.

Different types of early adversity including physical, emotional, sexual abuse, or psychosocial deprivation are

associated with altered DNA methylation in specific sites of the epigenome. The epigenome-wide studies that include a greater number of genetic loci, to date, are scarce, the sample sizes are small, assess early adverse events with different methods (Yang et al., 2013; Cecil et al., 2016; Kumsta et al., 2016; Perna et al., 2020; and Merrill et al., 2021) and their findings, are still inconsistent among them and with candidate genes studies.

Epigenetics changes and psychotherapy in BPD

A limited number of studies have explored the effect of psychotherapy on epigenetic changes in BPD.

One study was performed on a sample of 115 outpatients diagnosed with BPD (and 52 healthy controls) and extracted DNA from blood leukocytes before and after 4 weeks of Intensive Dialectical Behavioral psychotherapy (DBT) to measure CpG methylation of exons I and IV of the *BDNF* gene. Patients who responded to DBT, exhibited a decrease in DNA methylation of *BDNF* gene exons I and IV, whereas no association was found between BPD diagnosis and methylation levels (Perroud et al., 2013). In another study performed on 44 patients with BPD and 44 matched controls, DNA methylation of *APBA3* and *MCF2* genes was measured from blood samples. *APBA3* (neuronal adapter protein) is related to the production of β -amyloid, a component of amyloid plaques linked to Alzheimer's disease and *MCF2* is a guanine nucleotide exchange factor, involved in neurite outgrowth that has been associated with schizophrenia and autism-spectrum disorders. Individuals with BPD who respond to DBT therapy presented higher methylation in both genes after 12 weeks relative to non-responders (Knoblich et al., 2018).

A third study, involving a sample of 41 individuals with BPD and 41 healthy controls and assessing candidate gene DNA methylation, reported higher methylation levels in promoter IV of the *BDNF* gene in both saliva and blood samples of BPD patients. Twenty-six out of the 41 patients completed DBT psychotherapy and after 12 weeks, a decrease in methylation levels was observed only in saliva samples (Thomas et al., 2018).

Some results are discordant, for example, Perroud et al.'s (2013) study found significant DNA methylation change after therapy in blood samples while Thomas et al.'s (2018) work did not. One possible explanation may be that the first study reported an average of methylation from exon IV while the second reported individual CpG methylation. There were also differences in the methylation evaluation technique (high resolution melt analysis vs. pyrosequencing).

These initial findings suggest that psychotherapy may be associated with epigenetic changes in candidate genes related to neuroplasticity.

FKBP5 gene, early stress, BPD, and psychotherapy response

FKBP5 gene encodes for FK506, a glucocorticoid receptor co-chaperon whose levels are increased after stress exposure and decreases the ability of the glucocorticoid receptor to bind cortisol and to translocate to the nucleus, creating an ultrashort negative feedback for NR3C1 activation (Binder, 2009; Zannas and Binder, 2014). DNA methylation of *FKBP5* gene promoting regions decrease gene transcription and might limit the effects over stress neuroendocrine response (Zannas and Binder, 2014).

An association has been described between a lower DNA methylation of intron 7 of the *FKBP5* gene and the presence of child maltreatment in adults (Klengel et al., 2013) and preschool children (Tyrka et al., 2015). Moreover, high *FKBP5* DNA methylation was found in infants who displayed resistant attachment behavior (Mulder et al., 2017).

Changes in *FKBP5* DNA methylation have also been associated with other early stressors such as low socioeconomic status in childhood (Needham et al., 2015) and Holocaust survivors and their offspring (Yehuda et al., 2016).

These findings suggest a possible role for *FKBP5* in the adaptation of molecular stress response systems in relation to the early environment through epigenetic modifications.

Interestingly, in a case-control study with individuals with BPD, *FKBP5* intron 7 (bin 2), DNA methylation inversely correlate with empathic perspective taking and with anxiety symptoms. Furthermore, lower empathic perspective-taking abilities and anxiety correlated with childhood maltreatment. Although this study does not find differences between clinical and non-clinical samples, it reveals the relationship between empathy engagement and *FKBP5* DNA methylation (Flasbeck and Brüne, 2021).

FKBP5 DNA methylation has been also associated with response to psychotherapy in PTSD (Yehuda et al., 2013), children with anxiety disorders (Roberts et al., 2015), and agoraphobia (Roberts et al., 2019). Bishop et al. (2018) also report significant findings in individuals with PTSD treated with Mindfulness-Based Stress Reduction (MBSR) therapy, but in the opposite direction, with responders having increased DNA methylation (intron 7, bin 2).

Mentalization in borderline personality disorder

Mentalization can be understood as a mental activity that allows interpreting behavior in terms of intentional mental states of others (needs, desires, feelings, goals), constituting a form of social cognition (Fonagy and Luyten, 2009).

Through mentalization individuals realize that they have a mind that can mediate its experience with the world, have the

capacity to distinguish the inner reality from external reality and include both, an intrapersonal mental context and emotional processes of interpersonal communication (Gergely et al., 2002).

The quality of early attachment relationships is critical for the development of mentalization, as they allow the internal states to be mirrored by an attentive and reliable caretaker. This process at the same time impacts the processes of emotional regulation and self-control (Fonagy and Luyten, 2009).

A mentalization deficiency occurs in subjects with BPD compared to a non-clinical sample and with subjects with other personality disorders, but only in the presence of child abuse (Fonagy et al., 1996). However, other studies show a superior capacity for mentalization in these patients. This may be because the expression of deficits could be in BPD, specifically activated in the context of attachment relationships under conditions of high emotional arousal (Antonsen et al., 2016), which increases the tendency to attribute mental states to others that exceed the information given by social cues (hypermentalization; Sharp and Fonagy, 2015).

Less certainty about mental states was reported measured by the Reflective Functioning Questionnaire (RFQ) in a sample of individuals with BPD compared to healthy controls (Morandotti et al., 2018). Moreover, individuals with BPD would not present difficulties in the processes of decoding mental states from observed behavior, but rather to make causal inferences and predictions that include context information and basic social knowledge (Németh et al., 2020). Lower values of mentalization ability measured with RFQ mediate the relationship between a diagnosis of BPD and insecure adult attachment, supporting the idea that the presence of negative internal work models reduces the ability to accurately distinguish the relationship between mental states and behavior (Badoud et al., 2018).

The treatment specifically developed to increase mentalization, Mentalization-based Treatment (MBT; Bateman and Fonagy, 2004) has shown to be effective compared to Structured Clinical Management in a sample of individuals with BPD in both reducing self-injurious behaviors and hospitalizations and in reducing symptoms and improving interpersonal functioning (Bateman and Fonagy, 2009). Also, at an 8-year follow-up, patients treated with MBT maintained a stable improvement over time compared with Treatment as Usual (TAU; Bateman and Fonagy, 2008).

Using the Psychotherapy Process Q-Set, an instrument that allows the rating of sessions according to how close they are to a prototypical session of their respective orientation, one study compared sessions of TFP, DBT, and therapy focused on mentalization. Interestingly, the prototype mentalization response correlated with all therapies, with a greater correlation on mentalizing the other (including the therapist) in TFP and more focused on the self in DBT (Goodman et al., 2015). These findings are in agreement with the statement that the development of mentalization corresponds to a common factor in BPD psychotherapies (Fonagy and Allison, 2014).

The capacity to navigate the interpersonal environment using the ability to mentalize can be improved by psychotherapies of different orientations in BPD. This change could be associated with biological changes at the epigenetic level specially in stress response systems.

Pilot study: epigenetic changes in psychotherapy of adolescents diagnosed with BPD

Methods

This study aims to explore changes in peripheral DNA methylation of *FKBP5* gene, which encodes for a stress response protein, in relation to psychotherapy, symptomatology, and underlying psychological processes in a sample of 11 female adolescents diagnosed with BPD. For this purpose, measures of early trauma, borderline and depressive symptoms, psychotherapy outcome, mentalization, and emotion regulation were studied longitudinally at baseline, 3 and 6 months. Percentage DNA methylation levels of specific regions of *FKBP5* gene intron 7 were measured at the same time interval. The design was a quasi-experimental, longitudinal, process-outcome study.

Participants

Participants were female adolescent patients, aged 15–20 years, with a BPD subthreshold cut-off of 3 or more criteria of the DSM IV-TR for BPD. Subthreshold-BPD was included based on impairment of quality of life, presence of self-injury, and suicidality similar to full-syndrome BPD female adolescents previously reported (Kaess et al., 2017). Participants were starting a psychotherapeutic process with a focus on difficulties in the development of their personality, eight of them were of psychodynamic orientation and three were dialectical behavior therapy (DBT).

The exclusion criteria were the following: psychosis, pervasive developmental disorder, and unstable medical (non-psychiatric) disease.

Recruitment

A convenience sampling technique was used by contacting psychotherapists who work with adolescent populations and whose theoretical model and therapeutic approach include the development of the mentalizing capacity, including psychodynamic psychotherapies and DBT applied in private practice and in public and private outpatient treatment centers

of Santiago de Chile. Therapists were clinical psychologists or psychiatrists with formal therapeutic training. Verbal and written information about the research and subject participation were provided. The study was approved by the ethics committee of Pontificia Universidad Católica de Chile.

Measures

Symptomatic profiling, process, and outcome questionnaires

Eligibility criteria was assessed through the Structured Clinical Interview for DSM-IV axis II Personality Disorders (SCID II; First et al., 1995) and Mini International Neuropsychiatric Interview (M.I.N.I.-Kid; Sheehan et al., 2010). The presence of childhood trauma was evaluated using the Childhood Trauma Questionnaire (CTQ; Bernstein et al., 1997), while attachment patterns were evaluated using the Attachment Adolescent Questionnaire (AAQ; West et al., 1998). Once selected, the Brief Reflective Functioning Interview (BRFI; Rudden et al., 2005) and Difficulties in Emotion Regulation Scale (DERS; Gratz and Roemer, 2004) were applied at 0, 3, and 6 months. The Reflective Functioning Scale (RFS) coding system was applied to BRFI transcript by a certified coder (Fonagy et al., 1998). Also, as outcome measures the Youth Outcome Self-Report (Y-OQ-SR; Wells et al., 2003), Beck Depression Inventory (BDI-I; Beck et al., 1961), and Borderline Symptom List (BSL-23; Bohus et al., 2009) were applied at 0, 3 months and 6 months.

DNA methylation

In addition, a blood sample was collected to perform the epigenetic analysis at 0, 3, and 6 months. For each participant, a nurse collected three samples of 5 ml of venous blood. The genomic DNA of the participants was extracted from 5 ml leukocytes of peripheral venous blood using tubes with EDTA as an anticoagulant.

The methylation status of three intron 7 CpG sites (ADS3828-FS2, ADS3828-FS1, and ADS6607-FS) was determined individually as an artificial C/T SNP using QCpG software (Pyrosequencing method, Qiagen). The methylation level at each CpG site was calculated as the percentage of the methylated alleles divided by the sum of all methylated and unmethylated alleles. The mean methylation level was calculated using methylation levels of all measured CpG sites within the targeted region of each gene¹.

¹ <https://www.epigenix.com/d/service/pyrosequencing/dna-methylation>

Data analysis

Data analysis was performed using R version 3.4.1 (R Core Team, 2016) and the R packages *psych* for descriptive data (Revelle, 2016). Baseline features were summarized using descriptive statistics. Given multiple points of data for each participant, a *mixed-effect growth curve modeling strategy* was followed to assess the change in time of DNA methylation and clinical parameters that account for shared variance within subjects while modeling between-subject differences using the nlme r package (Pinheiro et al., 2013).

Our model included average FKBP5 DNA methylation as the dependent variable, predicted by a linear function of time (i.e., Time = 0, 1, 2), dummy coded presence or absence of childhood trauma [based on at least one above-threshold score on the CTQ (Tyrka et al., 2015)], and genotype (where 1 means presence of T allele). Psychotherapy response was measured by the reliable change index (RCI) of the Y-OQ-SR. RCI was calculated using pre and post treatment score, standard deviation, and Cronbach alpha from the normative sample of the scale according to Jacobson and Truax formula (Jacobson and Truax, 1991; Bauer et al., 2004). The RCI allows to identify if the differences are greater than expected due to random error. An RCI that is greater than 1.96 correspond to the 97.5th percentile of a normal distribution and is equivalent to a statistically significant change ($p < 0.05$; Jacobson and Truax, 1991). The dependent variable was transformed to its natural logarithm to normalize the distribution of the residues. Visual inspections were performed to check assumptions of heteroscedasticity and normality of residues. Missing values (timepoints) were programmed to be omitted from each of the equations. Only random intercepts were accounted for, to avoid convergence problems due to the small sample size and number of parameters in the model.

Results

The study was planned with 34 individuals based on a power calculation to detect a 10 percent difference in DNA methylation. Sampling was restricted by the onset of the COVID-19 pandemic. Two individuals dropped out of the study at baseline, and there were no significant differences in the number of symptoms. A total of 11 individuals completed the study. The mean age was 16.77 ± 1.64 years. Nine of the 11 patients fulfill the threshold of 5 BPD criteria and two patients fulfill the sub-threshold criteria with four symptoms. All patients were above the cut-off score for the presence of depression (BDI-I) of 13 for Chilean population (Valdés et al., 2017) and the mean was in the range of severe depressive symptomatology (Beck et al., 1988).

The presence of childhood trauma was determined if at least one of the subscales of the CTQ scored above the threshold

TABLE 1 Regression analyses of mean FKBP5 DNA methylation change in time according to trauma and psychotherapy response.

	β (SE)	t
Parameter		
Fixed Effects		
Intercept	75.03 (1.95)**	38.45
Time	−0.48 (0.68)	−0.70
Trauma	2.27 (2.22)	1.02
Psychotherapy Response	4.24 (2.55)	1.65
Genotype	1.73 (1.47)	1.17
Time $\times$ Trauma	−0.14 (0.89)	−0.02
Trauma $\times$ Psychotherapy Response	1.2 (0.96)	−0.45
Time $\times$ Trauma $\times$ Psychotherapy Response	−3.18 (1.24)*	−2.57
Random Effects		
Intercept	2.12	
Residual	1.46	

Note: Estimates and standard errors (in parentheses). * $p < 0.05$, ** $p < 0.01$.

for moderate trauma. Seven participants (63.6%) had scores in the “moderate to severe” trauma range on at least one of the subscales. No significant differences in DNA methylation were found at baseline between individuals with and without the presence of childhood trauma (0.79 vs. 0.78, $p = 0.65$).

Response to psychotherapy as measured by Y-OQ-SR and according to the Reliable Change Index was associated with a decrease in mean FKBP5 DNA methylation only in those participants who reported the presence of moderate to severe childhood trauma ($\beta = -3.18$, $SE = 1.24$, $p = 0.04$; Table 1). Fixed effects explain 0.42 of the variance (marginal R^2 : 0.42, conditional R^2 : 0.81).

No change was observed over time in levels of mentalization, nor was there any significant association with changes in DNA methylation.

No significant relationship was observed between genotype, depressive symptoms, borderline symptoms, emotional regulation, and change in DNA methylation over time.

Discussion

In this study, a reduction in FKBP5 DNA methylation was observed in responders to therapy and especially in the group with early trauma.

The finding of decreased FKBP5 DNA methylation associated with response to psychotherapy replicates the results of previous studies in individuals with PTSD who were treated with exposure therapy (Yehuda et al., 2013), children with anxiety disorders treated with cognitive behavioral therapy (Roberts et al., 2015) and individuals with Agoraphobia with or without panic disorder (Roberts et al., 2019). Bishop et al. (2018) also report significant findings in individuals with PTSD treated with Mindfulness-Based Stress Reduction (MBSR) therapy, but in the opposite direction, responders have increased DNA methylation (intron 7, bin 2). This study found a decrease in FKBP5 DNA methylation in BPD phenotype and with different

types of psychotherapy (psychodynamic psychotherapy and dialectic behavior therapy) suggesting that psychotherapies, in general, can act as “environmental regulators” (Yehuda et al., 2013) through modification of expression of HPA-axis related genes across several mental disorders. DBT psychotherapy has previously been associated with DNA methylation change in other genes in individuals with BPD, but not with *FKBP5* (Perroud et al., 2013; Knoblich et al., 2018; Thomas et al., 2018).

The identification of plasticity genes can contribute to the advance in the identification of molecular markers of stable improvement in BPD. Interesting candidates are genes associated with the HPA axis, *NR3C1*, and *FKBP5*. Both showed stability in methylation for 2 years in healthy adults suggesting that it may be markers of stable changes and individual differences in stress response regulation (Di Sante et al., 2018). More research is needed to determine the patterns of variability and stability in time of methylation patterns of different genes, in different developmental periods and in clinical samples.

A striking finding of this study is that only those who reported the presence of early trauma exhibited a decrease in DNA methylation. This suggests that in this group the biological mechanism for developing BPD could be different than the no early trauma group.

Although in this study no difference in DNA methylation of *FKBP5* intron 7 was found at baseline and no effect of *FKBP5* SNP1360780 risk T allele, several studies have reported a relationship between the presence of childhood trauma and decreased DNA methylation levels in this region across different populations, preschool children, low-income adult population, Holocaust offspring, postpartum women, individuals with MDD, and individuals with psychotic disorders (Klengel et al., 2013; Yehuda et al., 2013; Tyrka et al., 2015; Tozzi et al., 2018; Grasso et al., 2020; Misiak et al., 2020) in particular those individuals carrying the *FKBP5* SNP1360780 risk T allele, suggesting the impact of early emotional environment on stress response systems and development of psychopathology throughout life.

No significant change in mentalization levels was observed, nor was an association found with response to psychotherapy or changes in DNA methylation. This is probably due to the difficulty of the instruments to detect changes in mentalizing capacity, which is highly context dependent. In individuals with BPD faced with interpersonal situations that generate emotional arousal, mentalizing capacity is deactivated and less sophisticated behavioral and emotional patterns are activated (Fonagy and Bateman, 2008). Instruments that can assess mentalization emerging from dyadic interaction such as psychotherapy sessions (Talia et al., 2019) could be more accurate and ecologically valid in finding episodes of mentalizing deactivations and mentalizing improvements across time.

In this study, only individuals who reported the presence of early trauma and who responded to psychotherapy exhibited

a decrease in DNA methylation. Other studies have reported positive associations between the presence of early adverse events and response to psychotherapy, adult individuals treated for chronic depression responded better to psychotherapy than to psychopharmacological treatment if they had a history of childhood abuse (Nemeroff et al., 2003). Similarly, adolescents with non-suicidal self-injury (NSSI) had a better response to psychotherapy in reducing the frequency of NSSI if they had reported adverse childhood experiences (Edinger et al., 2020). This differential response to the presence of early trauma is suggestive of specific mechanisms not only of symptomatology development but also of distinct mechanisms of recovery. In this sample, the differential response to psychotherapy at the level of DNA methylation may imply that some individuals are more permeable at the molecular level to both negative (early trauma) and positive influences (psychotherapy) from their affective environment.

According to the differential sensitivity model, individuals carrying “plastic alleles” who, faced with an early sub-optimal affective environment, would be more susceptible to develop psychopathology but can be also more susceptible to respond to positive social environments (Hammen et al., 2015). Psychotherapeutic interventions, understood as a factor capable of modifying the relationship with the current social environment, may have a greater effect on individuals who carry plastic alleles (Leighton et al., 2017; Jiménez et al., 2018).

In accordance with the above, a GWAS study of twins reported a polygenic score based on differences in sensitivity to develop anxiety disorders according to positive or negative parenting. In a second sample, individuals with major differential sensitivity polygenic score responded better to individual cognitive behavioral therapy (Keers et al., 2016). These results suggest that those individuals who present a greater sensitivity to the environment present more emotional problems if they experienced negative parenting, but they will also be the ones who will benefit most from more intensive forms of psychotherapy (Choi-Kain et al., 2017).

Close human relationships regulate optimal stimulation and modulate arousal levels and attenuate stress in order to improve the adaptation to the social environment. This phenomenon has been called “psychobiological attunement”, and has been explored in mother-child dyads and peer relations and can be observed from its behavioral, physiological, and biochemical correlates (Field, 2012). For example, intrusive mothers can upregulate infants’ developing stress systems, increasing cortisol levels in saliva (Tarullo et al., 2017).

During the establishment of the therapeutic relationship, the formation of an alliance between patient and therapist can lead to the restoration of “epistemic trust”, that is, to restore an individual’s confidence in obtaining from another human being knowledge relevant to his or her adaptation to the social world (Fonagy and Allison, 2014). This would be particularly relevant with individuals with BPD, in whom insecure attachment

patterns developed in sub-optimal interaction with their caregivers would involve chronic epistemic mistrust, i.e., a deficiency in the trustworthiness and relevance of interpersonal communication with the concomitant development of deficient behavioral and emotional patterns for establishing cooperation and the ability to repair ruptures in relationships with others, increase their mentalizing capacities and improve their adaptation to their social environment (Fonagy et al., 2015; Orme et al., 2019).

Psychotherapy as a special form of human relationship would then be constituted as a biologically embedded experience, capable of altering biological functions in a stable and long-term manner (Demetriou et al., 2015). Psychotherapy is constituted as a disrupter of the “external social recursion” that goes from the social environment to the neural systems, modifying the subjective perception of the interpersonal environment. At the same time, it is capable of changing the “internal physiologic recursion” that ranges from the Central Nervous System to gene expression, that includes hormonal systems, inflammatory molecules, and intracellular signal transduction (Slavich and Cole, 2013).

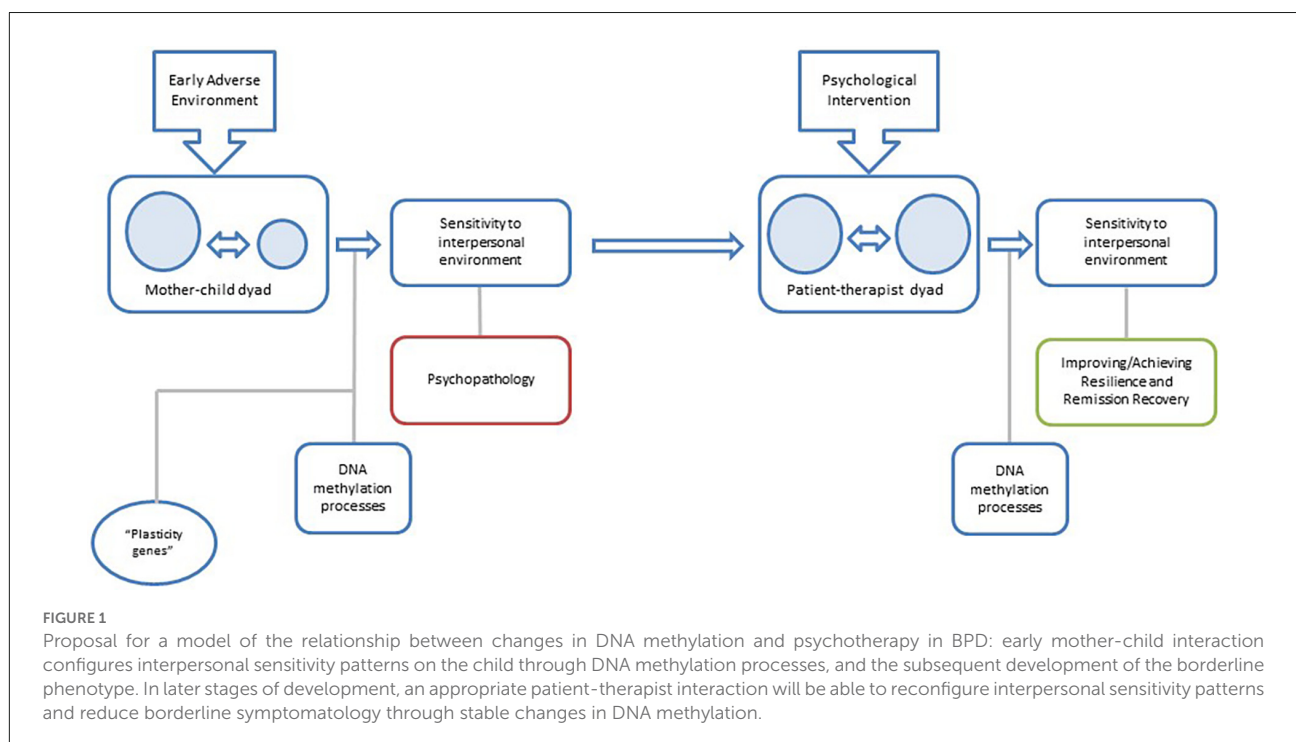
Psychotherapy focused on personality pathology may lead to changes in DNA methylation causing not only a symptomatic improvement but a reprogramming of the phenotypic adaptation to the interpersonal environment (Figure 1).

This study has several limitations; as a pilot study, our sample size was small relative to the number of predictors used in the final model, a known factor associated with type I and II errors. As such, our findings should be taken with caution and

not be interpreted as conclusive but as a valuable indication of the direction for further inquiry in the study of the molecular changes associated with psychotherapy for personality problems, as they aim to inspire further studies with an adequate sample size. For the same reason, other concomitant environmental factors were not incorporated as covariates that may interact with DNA methylation such as physical factors, i.e., nutrition, alcohol, drugs, contraceptives, and sleep deprivation (Nilsson et al., 2016; Sarabi et al., 2017; Gabbianelli and Damiani, 2018) and other social factors such as socioeconomic status (Maddock et al., 2018) should be taken into account.

The absence of healthy controls is another important limitation because in the childhood and adolescent population there may be methylation changes associated with development.

In this regard, studies exploring longitudinal changes using a genome-wide DNA methylation strategy in adolescents show that in a range of 3–6 months, there is one group of genes that is highly variable over time and another that varies between individuals, but remains stable over time (Lévesque et al., 2014). Moreover, a study in 51 adult individuals showed stability of *FKBP5* DNA methylation for 2 years, suggesting that it could be a trait marker of stable changes and individual differences in stress response regulation (Di Sante et al., 2018). The present study was able to compare and found differences between responders and non-responders to psychotherapy i.e., those individuals who had no significant clinical change over time operate as controls, in a manner similar to other studies that longitudinally explored *FKBP5* DNA methylation changes in relation to psychotherapy (Yehuda et al., 2013; Bishop et al., 2018).



In this study, BPD symptomatology was assessed through an instrument based on symptom intensity according to the DSM-IV categorization (Bohus et al., 2009). The study of personality can be broadened by resorting to a dimensional approach in line with the Alternative Model of Personality Disorders of DSM-5 (American Psychiatric Association, 2013) and the International Classification of Diseases, 11th Revision [ICD-11, (ICD-11—Mortality and Morbidity Statistics, 2021)], for example characterizing the Functioning Levels of Personality, which include identity, self-direction, empathy, and intimacy (Zimmermann et al., 2012). These dimensions of personality functioning are more directly connected with the focus of the therapeutic work and, therefore, allow a greater understanding of the mechanisms of change in psychotherapy of personality.

Along with outcome indicators, psychotherapy process measures such as emotional regulation and mentalization were included which has not been reported in previous studies. Other process measures can be incorporated, such as therapeutic alliance (Horvath et al., 2011), characteristics of the therapist (e.g., warmth), and the patient (e.g., expectations; Wampold, 2015). Prospective, longitudinal studies designs with the use of repeated measures could allow exploration of the interaction between DNA methylation changes and different factors of the therapy process.

Differences in the direction of DNA methylation change may be due to different regions of FKBP5 having been explored, for example, the region near promoter exon 1 (Yang et al., 2021) or intron 7 different CpG sites (Roberts et al., 2019; Bishop et al., 2018). Potentially, a reduction in methylation could generate a downstream lower expression of the FK506-binding protein 5 and reduce glucocorticoid receptor resistance and the impairment of negative feedback loop (Yang et al., 2021), however, functional inferences are not possible given that FKBP5 expression and endocrine markers of hypothalamic-pituitary-adrenal axis function were not measured. Yehuda et al. (2013) found that variation in FKBP5 DNA methylation was associated with treatment response and correlated with measures of plasma cortisol and glucocorticoid sensitivity, implying a functional impact at the HPA axis level of changes in DNA methylation.

Despite the limitations, the present work proposes a design that allows us to explore the explanatory relationships between the therapeutic process and changes at the epigenetic level, in other words, to advance in the understanding of the molecular mechanisms of psychic change.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committee of Pontificia Universidad Católica de Chile. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

YQ, LB, LH, and JJ contributed to the conception and design of the study. CH contributed to data analysis. YQ wrote the first draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by The National Fund for Science and Technology (Grant N° 1150166), The National Agency of Research and Development (Grant N° 21150647), and by the Fund for Innovation and Competitiveness (FIC) of the Chilean Ministry of Economy, Development and Tourism, through the Millennium Scientific Initiative, Grant N° IS130005.

Acknowledgments

We thank The National Agency of Research and Development, the Millenium Institute for Depression and Personality Research (MIDAP), Sussanne Schlüter-Müller MD, Klaus Schmeck, MD, and Ronan Zimmermann, PhD from the Personality Research and Research Department | Child and Adolescent Psychiatric Hospital, Psychiatric Hospitals of the University of Basel (UPK).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Amad, A., Ramoz, N., Thomas, P., Jardri, R., and Gorwood, P. (2014). Genetics of borderline personality disorder: systematic review and proposal of an integrative model. *Neurosci. Biobehav. Rev.* 40, 6–19. doi: 10.1016/j.neubiorev.2014.01.003
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596
- Antonsen, B. T., Johansen, M. S., Ro, F. G., Kvarstein, E. H., and Wilberg, T. (2016). Is reflective functioning associated with clinical symptoms and long-term course in patients with personality disorders? *Compr. Psychiatry* 64, 46–58. doi: 10.1016/j.comppsy.2015.05.016
- Badoud, D., Prada, P., Nicastro, R., Germond, C., Luyten, P., Perroud, N., et al. (2018). Attachment and reflective functioning in women with borderline personality disorder. *J. Pers. Disord.* 32, 17–30. doi: 10.1521/pedi.2017.31.283
- Bateman, A. W., and Fonagy, P. (2004). Mentalization-based treatment of BPD. *J. Pers. Disord.* 18, 36–51. doi: 10.1521/pedi.18.1.36.32772
- Bateman, A., and Fonagy, P. (2008). 8-year follow-up of patients treated for borderline personality disorder: mentalization-based treatment versus treatment as usual. *Am. J. Psychiatry* 165, 631–638. doi: 10.1176/appi.ajp.2007.07040636
- Bateman, A., and Fonagy, P. (2009). Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. *Am. J. Psychiatry* 166, 1355–1364. doi: 10.1176/appi.ajp.2009.09040539
- Bauer, S., Lambert, M. J., and Nielsen, S. L. (2004). Clinical significance methods: a comparison of statistical techniques. *J. Pers. Assess.* 82, 60–70. doi: 10.1207/s15327752jpa8201_11
- Beck, A. T., Steer, R. A., and Garbin, M. G. (1988). Psychometric properties of the beck depression inventory: twenty-five years of evaluation. *Clin. Psychol. Rev.* 8, 77–100. doi: 10.1016/0272-7358(88)90050-5
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., and Erbaugh, J. (1961). An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561–571. doi: 10.1001/archpsyc.1961.01710120031004
- Bernstein, D. P., Ahluvalia, T., Pogge, D., and Handelsman, L. (1997). Validity of the Childhood Trauma Questionnaire in an adolescent psychiatric population. *J. Am. Acad. Child Adolesc. Psychiatry* 36, 340–348. doi: 10.1097/00004583-199703000-00012
- Binder, E. B. (2009). The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. *Psychoneuroendocrinology* 34, S186–S195. doi: 10.1016/j.psyneuen.2009.05.021
- Bishop, J. R., Lee, A. M., Mills, L. J., Thurs, P. D., Eum, S., Clancy, D., et al. (2018). Methylation of FKBP5 and SLC6A4 in relation to treatment response to mindfulness based stress reduction for posttraumatic stress disorder. *Front. Psychiatry* 9:418. doi: 10.3389/fpsyt.2018.00418
- Bohus, M., Kleindienst, N., Limberger, M. F., Stieglitz, R. D., Domsalla, M., Chapman, A. L., et al. (2009). The short version of the Borderline Symptom List (BSL-23): development and initial data on psychometric properties. *Psychopathology* 42, 32–39. doi: 10.1159/000173701
- Bulbena-Cabre, A., Bassir Nia, A., and Perez-Rodriguez, M. M. (2018). Current knowledge on gene-environment interactions in personality disorders: an update. *Curr. Psychiatry Rep.* 20:74. doi: 10.1007/s11920-018-0934-7
- Calati, R., Gressier, F., Balestri, M., and Serretti, A. (2013). Genetic modulation of borderline personality disorder: systematic review and meta-analysis. *J. Psychiatr. Res.* 47, 1275–1287. doi: 10.1016/j.jpsychires.2013.06.002
- Cecil, C. A. M., Smith, R. G., Walton, E., Mill, J., McCrory, E. J., and Viding, E. (2016). Epigenetic signatures of childhood abuse and neglect: implications for psychiatric vulnerability. *J. Psychiatr. Res.* 83, 184–194. doi: 10.1016/j.jpsychires.2016.09.010
- Chanen, A. M., Jovev, M., Djaja, D., McDougall, E., Yuen, H. P., Rawlings, D., et al. (2008). Screening for borderline personality disorder in outpatient youth. *J. Pers. Disord.* 22, 353–364. doi: 10.1521/pedi.2008.22.4.353
- Choi-Kain, L. W., Finch, E. F., Masland, S. R., Jenkins, J. A., and Unruh, B. T. (2017). What works in the treatment of borderline personality disorder. *Curr. Behav. Neurosci. Rep.* 4, 21–30. doi: 10.1007/s40473-017-0103-z
- Demetriou, C. A., van Veldhoven, K., Relton, C., Stringhini, S., Kyriacou, K., and Vineis, P. (2015). Biological embedding of early-life exposures and disease risk in humans: a role for DNA methylation. *Eur. J. Clin. Invest.* 45, 303–332. doi: 10.1111/eci.12406
- Di Sante, J., Ismaylova, E., Nemoda, Z., Gouin, J. P., Yu, W. J., Caldwell, W., et al. (2018). Peripheral DNA methylation of HPA axis-related genes in humans: cross-tissue convergence, two-year stability and behavioural and neural correlates. *Psychoneuroendocrinology* 97, 196–205. doi: 10.1016/j.psyneuen.2018.07.019
- Ecker, S., Pancaldi, V., Valencia, A., Beck, S., and Paul, D. S. (2018). Epigenetic and transcriptional variability shape phenotypic plasticity. *Bioessays* 40:1700148. doi: 10.1002/bies.201700148
- Edinger, A., Fischer-Waldschmidt, G., Parzer, P., Brunner, R., Resch, F., and Kaess, M. (2020). The impact of adverse childhood experiences on therapy outcome in adolescents engaging in nonsuicidal self-injury. *Front. Psychiatry* 11:505661. doi: 10.3389/fpsyt.2020.505661
- Field, T. (2012). Relationships as regulators. *Psychology* 3, 467–479. doi: 10.4236/psych.2012.36066
- First, B., Spitzer, R., Gibbon, M., Williams, J., Davies, M., Borus, J., et al. (1995). The structured clinical interview for DSM-III-R personality disorders (SCID-II). Part II: multi-site test-retest reliability study. *J. Pers. Disord.* 9, 92–104. doi: 10.1521/pedi.1995.9.2.92
- Flasbeck, V., and Brüne, M. (2021). Association between childhood maltreatment, psychopathology and DNA methylation of genes involved in stress regulation: evidence from a study in Borderline Personality Disorder. *PLoS One* 16:e0248514. doi: 10.1371/journal.pone.0248514
- Fonagy, P., and Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy (Chic)* 51, 372–380. doi: 10.1037/a0036505
- Fonagy, P., and Bateman, A. (2008). The development of borderline personality disorder—a mentalizing model. *J. Pers. Disord.* 22, 4–21. doi: 10.1521/pedi.2008.22.1.4
- Fonagy, P., and Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Dev. Psychopathol.* 21, 1355–1381. doi: 10.1017/S0954579409990198
- Fonagy, P., Leigh, T., Steele, M., Steele, H., Kennedy, T., Mattoon, G., et al. (1996). The relation of attachment status, psychiatric classification and the response to psychotherapy. *J. Consult. Clin. Psychol.* 64, 22–31. doi: 10.1037/0022-006x.64.1.22
- Fonagy, P., Luyten, P., and Allison, E. (2015). Epistemic Petrification and the restoration of epistemic trust: a new conceptualization of borderline personality disorder and its psychosocial treatment. *J. Pers. Disord.* 29, 575–609. doi: 10.1521/pedi.2015.29.5.575
- Fonagy, P., Steele, M., Steele, H., and Target, M. (1998). *Reflective-Functioning Manual for Application to Adult Attachment Interviews (Ver 5)*. London: University College London.
- Gabbianelli, R., and Damiani, E. (2018). Epigenetics and neurodegeneration: role of early-life nutrition. *J. Nutr. Biochem.* 57, 1–13. doi: 10.1016/j.jnutbio.2018.01.014
- Gergely, G., Jurist, E. L., Fonagy, P. (2002). *Affect Regulation, Mentalization and the Development of the Self (1st ed.)*. London: Routledge.
- Goodman, G., Edwards, K., and Chung, H. (2015). The relation between prototypical processes and psychological distress in psychodynamic therapy of five inpatients with borderline personality disorder. *Clin. Psychol. Psychother.* 22, 83–95. doi: 10.1002/cpp.1875
- Grasso, D. J., Drury, S., Briggs-Gowan, M., Johnson, A., Ford, J., Lapidus, G., et al. (2020). Adverse childhood experiences, posttraumatic stress and FKBP5 methylation patterns in postpartum women and their newborn infants. *Psychoneuroendocrinology* 114:104604. doi: 10.1016/j.psyneuen.2020.104604
- Gratz, K. L., and Roemer, L. (2004). Multidimensional Assessment of emotion regulation and dysregulation: development, factor structure and initial validation of the difficulties in emotion regulation scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1023/B:JOBA.0000007455.08539.94
- Gunderson, J. G., and Lyons-Ruth, K. (2008). BPD's interpersonal hypersensitivity phenotype: a gene-environment-developmental model. *J. Pers. Disord.* 22, 22–41. doi: 10.1521/pedi.2008.22.1.22
- Hammen, C., Bower, J. E., and Cole, S. W. (2015). Oxytocin receptor gene variation and differential susceptibility to family environment in predicting youth borderline symptoms. *J. Pers. Disord.* 29, 177–192. doi: 10.1521/pedi.2014.28.152

- Horvath, A. O., Del Re, A. C., Flückiger, C., and Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy (Chic)* 48, 9–16. doi: 10.1037/a0022186
- ICD-11—Mortality and Morbidity Statistics (2021). Available online at: <https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fid.who.int%2fcd%2fentfity%2f941859884>
- Jacobson, N. S., and Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Consult. Clin. Psychol.* 59, 12–19. doi: 10.1037//0022-006x.59.1.12
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33, 245–254. doi: 10.1038/ng1089
- Jiménez, J. P., Botto, A., Herrera, L., Leighton, C., Rossi, J. L., Quevedo, Y., et al. (2018). Psychotherapy and genetic neuroscience: an emerging dialog. *Front. Genet.* 9:257. doi: 10.3389/fgene.2018.00257
- Kaess, M., Fischer-Waldschmidt, G., Resch, F., and Koenig, J. (2017). Health related quality of life and psychopathological distress in risk taking and self-harming adolescents with full-syndrome, subthreshold and without borderline personality disorder: rethinking the clinical cut-off? *Borderline Personal. Disord. Emot. Dysregul.* 4:7. doi: 10.1186/s40479-017-0058-4
- Keers, R., Coleman, J. R. I., Lester, K. J., Roberts, S., Breen, G., Thastum, M., et al. (2016). A genome-wide test of the differential susceptibility hypothesis reveals a genetic predictor of differential response to psychological treatments for child anxiety disorders. *Psychother. Psychosom.* 85, 146–158. doi: 10.1159/000444023
- Klengel, T., Mehta, D., Anacker, C., Rex-Haffner, M., Pruessner, J. C., Pariante, C. M., et al. (2013). Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat. Neurosci.* 16, 33–41. doi: 10.1038/nn.3275
- Knoblich, N., Gundel, F., Brückmann, C., Becker-Sadzio, J., Frischholz, C., and Nieratschker, V. (2018). DNA methylation of APBA3 and MCF2 in borderline personality disorder: potential biomarkers for response to psychotherapy. *Eur. Neuropsychopharmacol.* 28, 252–263. doi: 10.1016/j.euroneuro.2017.12.010
- Kumsta, R., Marzi, S. J., Viana, J., Dempster, E. L., Crawford, B., Rutter, M., et al. (2016). Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of CYP2E1. *Transl. Psychiatry* 6:e830. doi: 10.1038/tp.2016.95
- Leichsenring, F., Leibing, E., Kruse, J., New, A. S., and Leweke, F. (2011). Borderline personality disorder. *Lancet* 377, 74–84. doi: 10.1016/S0140-6736(10)61422-5
- Leighton, C., Botto, A., Silva, J. R., Jiménez, J. P., and Luyten, P. (2017). Vulnerability or sensitivity to the environment? Methodological issues, trends and recommendations in gene-environment interactions research in human behavior. *Front. Psychiatry* 8:106. doi: 10.3389/fpsy.2017.00106
- Lévesque, M. L., Casey, K. F., Szyf, M., Ismaylova, E., Ly, V., Verner, M. P., et al. (2014). Genome-wide DNA methylation variability in adolescent monozygotic twins followed since birth. *Epigenetics* 9, 1410–1421. doi: 10.4161/15592294.2014.970060
- Lutz, P. E., and Turecki, G. (2014). DNA methylation and childhood maltreatment: from animal models to human studies. *Neuroscience* 264, 142–156. doi: 10.1016/j.neuroscience.2013.07.069
- Maddock, J., Wulaningsih, W., Fernandez, J. C., Ploubidis, G. B., Goodman, A., Bell, J., et al. (2018). Associations between body size, nutrition and socioeconomic position in early life and the epigenome: a systematic review. *PLoS One* 13:e0201672. doi: 10.1371/journal.pone.0201672
- Martín-Blanco, A., Ferrer, M., Soler, J., Salazar, J., Vega, D., Andiñ, O., et al. (2014). Association between methylation of the glucocorticoid receptor gene, childhood maltreatment and clinical severity in borderline personality disorder. *J. Psychiatr. Res.* 57, 34–40. doi: 10.1016/j.jpsychires.2014.06.011
- Merrill, S. M., Gladish, N., Fu, M. P., Moore, S. R., Konwar, C., Giesbrecht, G. F., et al. (2021). Associations of peripheral blood DNA methylation and estimated monocyte proportion differences during infancy with toddler attachment style. *Attach. Hum. Dev.* 1, 1–30. doi: 10.1080/14616734.2021.1938872. [Online ahead of print].
- Misiak, B., Karpiński, P., Szmida, E., Grądzewski, T., Jabłoński, M., Cyranka, K., et al. (2020). Adverse childhood experiences and methylation of the FKBP5 gene in patients with psychotic disorders. *J. Clin. Med.* 9:E3792. doi: 10.3390/jcm9123792
- Morandotti, N., Brondino, N., Merelli, A., Boldrini, A., De Vidovich, G. Z., Ricciardo, S., et al. (2018). The Italian version of the Reflective Functioning Questionnaire: validity data for adults and its association with severity of borderline personality disorder. *PLoS One* 13:e0206433. doi: 10.1371/journal.pone.0206433
- Mulder, R. H., Rijlaarsdam, J., Luijk, M. P. C. M., Verhulst, F. C., Felix, J. F., Tiemeier, H., et al. (2017). Methylation matters: FKBP5 binding protein 51 (FKBP5) methylation moderates the associations of FKBP5 genotype and resistant attachment with stress regulation. *Dev. Psychopathol.* 29, 491–503. doi: 10.1017/S095457941700013X
- Murgatroyd, C., Patchev, A. V., Wu, Y., Mical, V., Bockmühl, Y., Fischer, D., et al. (2009). Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat. Neurosci.* 12, 1559–1566. doi: 10.1038/nn.2436
- Needham, B. L., Smith, J. A., Zhao, W., Wang, X., Mukherjee, B., Kardia, S. L., et al. (2015). Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. *Epigenetics* 10, 958–969. doi: 10.1080/15592294.2015.1085139
- Nemeroff, C. B., Heim, C. M., Thase, M. E., Klein, D. N., Rush, A. J., Schatzberg, A. F., et al. (2003). Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proc. Natl. Acad. Sci. U S A* 100, 14293–14296. doi: 10.1073/pnas.2336126100
- Németh, N., Péterfalvi, Á., Czéh, B., Tényi, T., and Simon, M. (2020). Examining the relationship between executive functions and mentalizing abilities of patients with borderline personality disorder. *Front. Psychol.* 11:1583. doi: 10.3389/fpsyg.2020.01583
- Nilsson, E. K., Bostrom, A. E., Mwinyi, J., and Schiöth, H. B. (2016). Epigenomics of total acute sleep deprivation in relation to genome-wide DNA methylation profiles and RNA expression. *OMICS* 20, 334–342. doi: 10.1089/omi.2016.0041
- Orme, W., Bowersox, L., Vanwoerden, S., Fonagy, P., and Sharp, C. (2019). The relation between epistemic trust and borderline pathology in an adolescent inpatient sample. *Borderline Personal. Disord. Emot. Dysregul.* 6:13. doi: 10.1186/s40479-019-0110-7
- Perna, L., Zhang, Y., Wild, B., Kliegel, M., Ihle, A., Schöttker, B., et al. (2020). Childhood exposure to hunger: associations with health outcomes in later life and epigenetic markers. *Epigenomics* 12, 1861–1870. doi: 10.2217/epi-2019-0333
- Perroud, N., Salzmann, A., Prada, P., Nicastro, R., Hoeppli, M. E., Furrer, S., et al. (2013). Response to psychotherapy in borderline personality disorder and methylation status of the BDNF gene. *Transl. Psychiatry* 3:e207. doi: 10.1038/tp.2012.140
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. R., and Development Core Team (2013). nlme: Linear and nonlinear mixed effects models. R Package Version 3, 1–108. Available online at: <https://CRAN.R-project.org/package=nlme>.
- Pyrosequencing method. Available online at: <https://www.epigenix.com/d/service/pyrosequencing/dna-methylation>.
- R Core Team (2016). R: a language and environment for statistical computing. Vienna: Austria. Available online at: <https://www.R-project.org/>.
- Radtke, K. M., Schauer, M., Gunter, H. M., Ruf-Leuschner, M., Sill, J., Meyer, A., et al. (2015). Epigenetic modifications of the glucocorticoid receptor gene are associated with the vulnerability to psychopathology in childhood maltreatment. *Transl. Psychiatry* 5:e571. doi: 10.1038/tp.2015.63
- Revelle, W. (2016). psych: Procedures for Personality and Psychological Research. Evanston, IL: Northwestern University. Available online at: <https://cran.r-project.org/web/packages/psych/psych.pdf>.
- Roberts, S., Keers, R., Breen, G., Coleman, J., Jöhren, P., Kepa, A., et al. (2019). DNA methylation of FKBP5 and response to exposure-based psychological therapy. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 180, 150–158. doi: 10.1002/ajmg.b.32650
- Roberts, S., Keers, R., Lester, K. J., Coleman, J. R., Breen, G., Arendt, K., et al. (2015). HPA axis related genes and response to psychological therapies: genetics and epigenetics. *Depress. Anxiety* 32, 861–870. doi: 10.1002/da.22430
- Rudden, M. G., Milrod, B., and Target, M. (2005). *The Brief Reflective Functioning Interview*. New York: Weill Cornell Medical College.
- Sarabi, M. M., Ghareghani, P., Khademi, F., and Zal, F. (2017). Oral contraceptive use may modulate global genomic DNA methylation and promoter methylation of APC1 and ESR1. *Asian Pac. J. Cancer Prev.* 18, 2361–2366. doi: 10.22034/APJCP.2017.18.9.2361
- Sharp, C., and Fonagy, P. (2015). Practitioner review: borderline personality disorder in adolescence - recent conceptualization, intervention and implications for clinical practice. *J. Child Psychol. Psychiatry* 56, 1266–1288. doi: 10.1111/jcpp.12449
- Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., et al. (2010). Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). *J. Clin. Psychiatry* 71, 313–326. doi: 10.4088/JCP.09m05305whi

- Skoglund, C., Tiger, A., Rück, C., Petrovic, P., Asherson, P., Hellner, C., et al. (2021). Familial risk and heritability of diagnosed borderline personality disorder: a register study of the Swedish population. *Mol. Psychiatry* 26, 999–1008. doi: 10.1038/s41380-019-0442-0
- Slavich, G. M., and Cole, S. W. (2013). The emerging field of human social genomics. *Clin. Psychol. Sci.* 1, 331–348. doi: 10.1177/2167702613478594
- Stead, V. E., Boylan, K., and Schmidt, L. A. (2019). Longitudinal associations between non-suicidal self-injury and borderline personality disorder in adolescents: a literature review. *Borderline Personal. Disord. Emot. Dysregul.* 6:3. doi: 10.1186/s40479-019-0100-9
- Talia, A., Miller-Bottomo, M., Katznelson, H., Pedersen, S. H., Steele, H., Schröder, P., et al. (2019). Mentalizing in the presence of another: measuring reflective functioning and attachment in the therapy process. *Psychother. Res.* 29, 652–665. doi: 10.1080/10503307.2017.1417651
- Tarullo, A. R., St. John, A. M., and Meyer, J. S. (2017). Chronic stress in the mother-infant dyad: maternal hair cortisol, infant salivary cortisol and interactional synchrony. *Infant. Behav. Dev.* 47, 92–102. doi: 10.1016/j.infbeh.2017.03.007
- Thomas, M., Knoblich, N., Wallisch, A., Glowacz, K., Becker-Sadzio, J., Gundel, F., et al. (2018). Increased BDNF methylation in saliva, but not blood, of patients with borderline personality disorder. *Clin. Epigenetics* 10:109. doi: 10.1186/s13148-018-0544-6
- Tozzi, L., Farrell, C., Booi, L., Doolin, K., Nemoda, Z., Szyf, M., et al. (2018). Epigenetic changes of FKBP5 as a link connecting genetic and environmental risk factors with structural and functional brain changes in major depression. *Neuropsychopharmacology* 43, 1138–1145. doi: 10.1038/npp.2017.290
- Trull, T. J., Jahng, S., Tomko, R. L., Wood, P. K., and Sher, K. J. (2010). Revised NESARC personality disorder diagnoses: gender, prevalence and comorbidity with substance dependence disorders. *J. Personal. Disord.* 24, 412–426. doi: 10.1521/pedi.2010.24.4.412
- Turecki, G., and Meaney, M. J. (2016). Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. *Biol. Psychiatry* 79, 87–96. doi: 10.1016/j.biopsych.2014.11.022
- Tyrka, A. R., Ridout, K. K., Parade, S. H., Paquette, A., Marsit, C. J., Seifer, R., et al. (2015). Childhood maltreatment and methylation of FKBP5 binding protein 5 gene (FKBP5). *Dev. Psychopathol.* 27, 1637–1645. doi: 10.1017/S0954579415000991
- Valdés, C., Morales-Reyes, I., Pérez, J. C., Medellín, A., Rojas, G., Krause, M., et al. (2017). Propiedades psicométricas del inventario de depresión de Beck IA para la población chilena (Psychometric properties of a spanish version of the Beck depression inventory IA). *Revista Médica de Chile* 145, 1005–1012. doi: 10.4067/s0034-98872017000801005
- Wampold, B. E. (2015). How important are the common factors in psychotherapy? an update. *World Psychiatry* 14, 270–277. doi: 10.1002/wps.20238
- Weaver, I. C. (2007). Epigenetic programming by maternal behavior and pharmacological intervention. Nature versus nurture: let's call the whole thing off. *Epigenetics* 2, 22–28. doi: 10.4161/epi.2.1.3881
- Wells, M. G., Burlingame, G. M., and Rose, P. M. (2003). Youth outcome questionnaire self report. Wilmington, DE: American Professional Credentialing Services.
- West, M., Rose, M. S., Spreng, S., Sheldon-Keller, A., and Adam, K. (1998). Adolescent attachment questionnaire: a brief assessment of attachment in adolescence. *J. Youth Adolesc.* 27, 661–673. doi: 10.1023/A:1022891225542
- Winograd, G., Cohen, P., and Chen, H. (2008). Adolescent borderline symptoms in the community: prognosis for functioning over 20 years. *J. Child Psychol. Psychiatry* 49, 933–941. doi: 10.1111/j.1469-7610.2008.01930.x
- Yang, B. Z., Zhang, H., Ge, W., Weder, N., Douglas-Palumberi, H., Perepletchikova, F., et al. (2013). Child abuse and epigenetic mechanisms of disease risk. *Am. J. Prev. Med.* 44, 101–107. doi: 10.1016/j.amepre.2012.10.012
- Yang, R., Xu, C., Bierer, L. M., Flory, J. D., Gautam, A., Bader, H. N., et al. (2021). Longitudinal genome-wide methylation study of PTSD treatment using prolonged exposure and hydrocortisone. *Transl. Psychiatry* 11:398. doi: 10.1038/s41398-021-01513-5
- Yehuda, R., Daskalakis, N. P., Bierer, L. M., Bader, H. N., Klengel, T., Holsboer, F., et al. (2016). Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biol. Psychiatry* 80, 372–380. doi: 10.1016/j.biopsych.2015.08.005
- Yehuda, R., Daskalakis, N. P., Desarnaud, F., Makotkine, I., Lehrner, A. L., Koch, E., et al. (2013). Epigenetic biomarkers as predictors and correlates of symptom improvement following psychotherapy in combat veterans with PTSD. *Front. Psychiatry* 4:118. doi: 10.3389/fpsy.2013.00118
- Zannas, A. S., and Binder, E. B. (2014). Gene-environment interactions at the FKBP5 locus: sensitive periods, mechanisms and pleiotropism. *Genes Brain Behav.* 13, 25–37. doi: 10.1111/gbb.12104
- Zannas, A. S., Wiechmann, T., Gassen, N. C., and Binder, E. B. (2015). Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology* 41, 261–274. doi: 10.1038/npp.2015.235
- Zimmermann, J., Ehrenthal, J. C., Cierpka, M., Schauenburg, H., Doering, S., and Benecke, C. (2012). Assessing the level of structural integration using operationalized psychodynamic diagnosis (OPD): implications for DSM-5. *J. Pers. Assess.* 94, 522–532. doi: 10.1080/00223891.2012.700664



OPEN ACCESS

EDITED BY

Mark Leonard Solms,
University of Cape Town, South Africa

REVIEWED BY

Matthew John Mellor,
Essex Partnership University NHS
Foundation Trust, United Kingdom
Jesse Proudfoot,
Durham University, United Kingdom
Jessica Tran The,
INSERM U1077 Neuropsychologie et
Imagerie de la Mémoire Humaine,
France

*CORRESPONDENCE

Gustaw Sikora
sikoragustaw@yahoo.co.uk

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 29 May 2022

ACCEPTED 04 October 2022

PUBLISHED 20 October 2022

CITATION

Sikora G (2022) An economic model
of the drives from Friston's free energy
perspective.
Front. Hum. Neurosci. 16:955903.
doi: 10.3389/fnhum.2022.955903

COPYRIGHT

© 2022 Sikora. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

An economic model of the drives from Friston's free energy perspective

Gustaw Sikora*

British Psychoanalytical Society, Institute of Psychoanalysis, London, United Kingdom

This paper is focused on the theory of drives, particularly on its economic model, which was an integral part of Freud's original formulation. Freud was aiming to make a link between the psychic energy of drives and the biophysical rules of nature. However, he was not able to develop this model into a comprehensive system linking the body and the mind. The further development of psychoanalytic theory, in various attempts to comprehend the theory of drives, can be described as taking different approaches. Some of them equate drives with bodily impulses, others abandon the economic model, a few stay with Freud's original model. I believe that the Friston notion of free energy and the hierarchical model of the brain allows us to develop this model and to integrate the economic model into some contemporary theories of drives. I argue against those theories equating drives with biological impulses. My arguments are supported by Freud's project itself but also by recent developments within neuro-psychoanalysis describing the process of mentalizing homeostasis, interoceptive signals and relations with caregivers. I argue for those theories which see the drives as psychic forces, which through developmental processes and cathexes acquire aims and objects, and become intertwined with impulses originating internally and externally, such as affect, interoceptive impulses, perception of the external world, and impulses from erotogenic zones in particular. Here, I present my analysis of the compatibility and consistency of free energy and the hierarchical model perspective, and two psychoanalytical traditions of thoughts: French psychoanalysis and the post-Kleinian school of British psychoanalysis. In particular, my analysis focuses on the contemporary Kleinian notion of unconscious phantasies, especially Bronstein's description of their semiotic aspects. Secondly, I look at Segal's view of drives as dialectic forces of adaptation vs. conservatism. Analyzing the French tradition, I focus on Green's perspective on the drives, Lacan's distinction between drives and desire, and Penot's description of the process of subjectivation. I conclude that free energy, as described by Friston, can be seen as a source of the drives' energy and the process of minimizing it is an equivalent of what Freud described as binding the free energy, in which psychic unbound energy acquires distinctive features and becomes bound.

KEYWORDS

economic model, theory of drives, free energy, somato-psychic frontier, psychoanalysis

Introduction

Psychoanalytic metapsychology is an attempt to build a coherent model of the mind. Although in this endeavor we make use of data from clinical observations and research, many parts of emerging models remain speculative (see [Britton, 2015](#)). In recent years, there have been many attempts to use discoveries from natural sciences to enhance and clarify psychoanalytic knowledge.

This paper focuses on a particular aspect of one of the psychoanalytic models, namely on the economic aspect of the theory of drives. It is based on assumptions which Freud made when he formulated his model. For clarity, in this paper I use the term “psychoanalysis” to describe models of the mind which use the theory of drives.

The theory of drives is one of the fundamental pillars of psychoanalysis. For Freud, the economic model was an integral part of this theory, which aimed to make the link between psychic energy and the biophysical rules of nature ([Freud, 1915a](#)). However, he was unable to develop this model into a comprehensive system linking body and mind. That has left psychoanalysis with many questions, one of them being: what is the nature of the energy of the drives? One might also consider this question from the perspective of the twofold nature of the phenomenon: the energy and the content of the drives (quantity and quality) and focusing on a qualitative aspect only. However, nature of energy still remains an open question.

In the various attempts to comprehend drive theory, most of the further developments of psychoanalytic theory can be described as following two different avenues. One equates drives with recognizable bodily impulses; for example, [Sandler and Schachter \(2014\)](#), [Solms \(2018\)](#). The other ones follow [Freud \(1933\)](#) in his claim that drives are “mythical entities” consisting of pure psychic energy; but they more or less abandon his economic model (see [Bell, 2015](#); [Penot, 2017](#)). There are exceptions to these: most notable [Green, 2010, 2015a,b](#), who follows Freud’s economic model without explaining the biological sources of the drives’ energy but not denying them either. I will not consider in this paper those models that abandon drives completely, as abandoning drives means a complete departure from the psychoanalytic model as it is understood in this paper.

This paper argues that Friston’s notion of free energy and the model of the brain as a “probability machine” ([Friston and Stephan, 2007](#); [Friston, 2010](#)) provide us with knowledge that allows us to integrate the economic model into the contemporary theory of drives, including object-relation theories. Using developments within neuro-psychoanalysis that describe the process of mentalizing homeostasis and interoceptive signals, this paper attempts to develop a hypothesis concerning the nature of a particular function of brain/mind which helps the free energy of the drives to become bound, and the drives to acquire content.

The paper also tries to match this hypothesis with those theories that see drives as psychic forces which, through developmental processes, become intertwined with impulses originating internally and externally, such as affect, interoceptive impulses, perception of the external world, and impulses from erotogenic zones in particular.

While seeking for compatibility and consistency of free energy, and a hierarchical predictive model perspective with traditions of psychoanalytical thought, I will use mostly French psychoanalysis and the post-Kleinian school, as I find that both are strongly rooted in Freud’s original formulation of the drives. This analysis is mostly focused on the contemporary Kleinian notion of unconscious phantasies as representations of the drives, in particular [Bronstein’s \(2015\)](#) description which uses Kristeva’s work on semiotics ([Kristeva, 1984, 2000](#)) and on [Green’s \(2015a\)](#) understanding of drives as forces of binding and unbinding (see also [Penot, 2017](#)).

There have already been many attempts to bridge neuroscience and psychoanalysis, all facing difficulties which come up from the complexities of both fields, different epistemologies, terminologies (they both use similar or the same terms to describe different phenomena), and many more. These often lead to misunderstanding and sometimes premature rejection of such attempts by psychoanalysts. On the other hand, they sometimes lead to the building of radical reformulations of psychoanalytic theory which are not compatible with clinical observations and practice. To minimize these problems, I will clarify meanings of analytic terms I am using as much as possible, yet being aware that some analysts use those terms differently. Also, despite my starting point being Freud’s formulation, I am focused on compatibility with contemporary psychoanalytic knowledge.

Ultimately, my analysis concludes that free energy, as described by [Friston and Stephan \(2007\)](#), [Friston \(2010\)](#), can be seen as a biophysical source of the drives, and that the process of minimizing this free energy is equivalent to what Freud described as a process of binding energy, in which purely unbound energy acquires distinctive features. It offers an answer to the question of what force is behind crucial processes of cathexis/deathexis. The paper concludes by drawing attention to some possible implications of this hypothesis for psychoanalytic metapsychology.

Theory of the drives

First, I need to seek as precise of the term drive as possible which Freud used in his model. There is a certain confusion about this term, especially in English speaking literature, mostly due to Strachey’s translation of Freud ([Strachey’s, 1957](#) [in introduction in [Freud, 1915b](#)], pp. 111–116). Significance of differentiation of a drive and an instinct was earlier elaborated by [Laplanche and Pontalis \(1973b\)](#). [Conrad \(2021\)](#),

to whose work I am indebted, believes that “the drives no longer occupy a preeminent position in the psychoanalytic theoretical landscape” because many analysts think that they are “simultaneously too abstract and too simple to be a serviceable concept.” Although, this may be true for some schools of psychoanalysis, there are also prominent analysts who have been of the opposite view, among them Segal (1985), Feldman (2000), Green (2015a), Bell (2017).

I start from the use of the term “drive” (original German term *Trieb*), prior to Freud’s writing. Conrad (2021, pp. 493–494) claims that a crucial figure in developing the meaning of this term was J. F. Blumenbach, who introduced the term “*Bildungstrieb*” (drive of development) to describe a force which pushes a living organism toward a determined shape; however, he saw it as “a mysterious force that could be clearly identified, while its cause remained unknown” (ibidem). I would add here that in contemporary language, “toward its determined shape” could be formulated as “toward organized structure.” Furthermore, drive (Blumenbach’s *Trieb*) can be seen as “anti-entropy” force, if we put it (drive) in the framework where entropy represents chaos and is the opposite of organized structure.

Further XIX century considerations on motivation brought about a better understanding of the complexity of it, including the works of Spencer (1870), Schneider (1880). Conrad (2021) summarizes it as follows: “a spectrum from animal to human as the progressive increase in complexity of motivation, with the simplest animal organisms experiencing feelings purely mechanistically, more complex animals being motivated perceptually, still more complex animals being motivated by representations, and finally the most complex animals (i.e., humans) being motivated by thoughts” (ibidem, p. 496). At the same time, Schopenhauer (1958) made instinct central to his understanding of human will. He believed that the ultimate force behind the complex phenomenology of human behavior is instinct, which is a purely biological force and is always hidden in various behaviors and actions, but the goal of it is always the same. Finally, we can look at Nietzsche’s (1881) work where he tried to elucidate the role of drives. He emphasized that the same situation (stimulus) can evoke very different responses from different individuals, depending on which drive is active at the very moment, emphasizing that different behavioral outcomes always aim to satisfy the drive involved. In other words, humans are driven by the same stimulus to satisfy different outcomes determined by the qualitative features of the drives involved.

I believe that Schopenhauer’s and Nietzsche’s takes on this problem elucidate a difference between an instinct and a drive in German speaking world of the time.

All of this was surely known to Freud when he started his own work on the formulation of drives. His scientific ambitions did not allow him to be satisfied with philosophical speculation; he wanted his theory to be proved by natural observations or other natural sciences (for more see Jones, 1953). Nevertheless,

Freud’s writings on drives leave considerable room for interpretation in attempts to define them. His own struggle with how to be precise in conceptualizing them can be seen in his words when, on the one hand, he described the theory of the drives as a fundamental concept [*grundbegriff*] (Freud, 1915b), while on the other he called them “mythical entities” (Freud, 1933). However, he drew clear lines that constitute his understanding of the drives and their central role in a psychoanalytic model of the mind. He elaborated on the qualitative dimension of them but in regard to their source, [Quelle] he was realistic about the limitations of his knowledge; he wrote: “The study of the sources of instincts (drives) lies outside the scope of psychology. . . in mental life we know them only by their aims” (Freud, 1915b).

I’ll summarize how Freud described the drives. Firstly, drives [*trieb*] are psychic forces that are distinguished from instincts. The latter are more biological in their nature, more automatic, and are less prone to modification or alteration. Drives are modified during development. Drives are described by four features: source, aim, object and pressure (Laplanche and Pontalis, 1973a,b). Aims and objects are changing and the outcome of such development varies (Freud, 1905). [for more arguments for differentiating instinct and drive see Conrad, 2021, p. 499–503].

Secondly, source and pressure, which also describe drives, remain the same (Freud, 1915b). The latter is conceived as a quantitative factor of the drives’ energy.

Thirdly, the source of the drives, identified as the Id, is “on the frontier between the mental and the physical” (Freud, 1905, 1915b). According to Freud, aim and object are attached by this energy and subsequently we have the “vicissitude” of the drive [*triebschicksal*], which is an entirely psychic entity. The nature of this energy and in what way it is attached to the aim and object remained unexplained. The part of the vicissitudes which bears energy, or in other words operates as a vehicle for already bound energy, is affect. Affects themselves have energy but the energy in question is one which links aim, object and affect in a one entity.

Fourthly, the drives have a dualistic nature (Freud, 1920); that dualism is the regulating force of the representatives (aims, objects) of the drives, which determines the drives’ vicissitudes (see also Bion, 1963).

Therefore, we can describe drive as a demand on the mind; that demand starts with stimuli and binds certain ideas with certain objects which become salient to the individual. Also, it is bound with the “quota of affect,” which determines the strength of the drive. This constitutes the qualitative features of the drives and gives them pressure coming from energetic aspect of affect. For Freud, there are two crucial elements of the drive: affects and ideas. Ideas describe aim and object, and affects are transpositions of the quantity of energy (Freud, 1915a,c; see also Laplanche and Pontalis, 1973a). It is very important to note that affects are not the same as drives. They may be bound with ideas or split off from them. They bear a quantitative dimension of

energy [*affektbetrag*], but they are part of the vicissitudes, not the source of the drives' initial energy (Freud, 1915a). Affects are basically conscious and subjective; Freud tried to make "affects" distinct from "quotas of affects," which more closely correspond to drives but are still not the same. "Quota" reflects the fact that affect's energy must also be bound (cathected), however, cathected is also strength of affect and that which bears the quantitative part of a drive and constitutes a pressure. Together with ideas (qualitative part) they constitute the vicissitudes of the drives. Conrad (2021) writes "[drive]...arouses an affective state that serves to evaluate the world in such a way that makes the salience of the object justified" and I would add to make object "desired." Therefore, we could say that the crucial process in forming drive's "content;" that is: aim, object and accompanying affect, is cathexis (*Besetzung*). Freud has never given a rigorous definition of this term; however, it is widely understood that it is a process which creates intentionality and a value object (Laplanche and Pontalis, 1973a). The opposite process of decathexis must also be active in the human mind so as to allow changes in the drive's aim and object. Again, that leaves us with the question: what force/s are behind cathexis and decathexis.

My interpretation of Freud's claim that "drives are mythical entities" is that we can use the psychoanalytic method to research qualitative features of the drives, which are attached to their energy and can be modified; but we have no access to the drives' sources of energy as such because they are "pure" energy. In the clinical sense, it means that psychoanalytic work with the unconscious is dealing with the drives' unconscious derivatives. Freud wrote "I am in fact of the opinion that antithesis of conscious and unconscious is not applicable to drives. Drive can never become an object of consciousness-only; the idea that represents it can. Even in the unconscious, moreover, a drive cannot be represented other than by an idea. If the drive did not attach itself to an idea, or manifest itself as an affective state, we could know nothing about it" (Freud, 1915a). Regarding the source, Freud postulated "continuously flowing source of stimulation," (Freud, 1905) or in other words a permanent demand on the mind coming from unknown source. I would suggest that Freud is referring here to source as continuing process which "produce" energy which demands [compare with definition of source (The Cambridge English Dictionary, 2022)].

The fact worthy of noting is that the number of drives is not limited and they can be shaped in a variety of ways. Making a list of them, or classifying them in nozological manner, seems to be impossible in this system. Therefore, Freud after a few attempts finally made a simple distinction between Life drives and Death (Anti-Life) drives (plural and various in their expressions). This distinction seems to be based on a practical/clinical criterion, which is the dimension: promoting development vs. restraining development (of the mind).

The further development of psychoanalysis had to deal with those considerable, visible gaps in the theory of the drives.

A complete analysis of that development is beyond my scope and is not the purpose of this paper. Also, such an analysis would be almost impossible because, as Sandler and Schachter (2014) argue, "every analyst has private theories and is influenced by different schools." Therefore, I outline only certain tendencies in that development; this allows me to explain my choices of models for further argument. Later, I focus on selected propositions which, in my view, have the theory of drives based on Freud's model described above; and also provide knowledge which is to some extent compatible with the free energy model. I am fully aware that not all analysts will agree with my outline and selection. Also, as it is only an outline, my description is simplified and does not give justice to all analytic thinkers. Most importantly, my selection here is dictated by the purpose of this paper, which is not to compare and analyze psychoanalytic schools of thoughts, but to develop a hypothesis that answers the question: What is the primal energy of the drives?

One line of that development was to "unhook the concept from the biological theorizing that formed part of Freud's initial presentation" (Bell, 2015). Bell is writing here about the death drive, but I believe he means the whole of the drive theory.

One of the proponents of this approach is the Kleinian school, which has developed a complex theory of the drives but has effectively abandoned the biological aspect and the economic model of the drives. The very question of the drives' origins is often replaced by the claim of innate purposive drives. Klein herself tended "to assert complementarity between her emerging viewpoint and Freud's theories" (Klein, 1940), but many of her followers believe that her work proved that the economic model was wrong. Spillius et al. (2011a) argue that Klein's understanding of the work of mourning in the context of infantile conflict of the depressive position "has more bite and also more depth than Freud's economic formulation of mourning, and incidentally shows the limitations of the economic model." However, showing limitations is not equal to disproving a claim. It seems to me that Klein has presented a greater complexity of the dynamics of mourning; however, this does not disprove the economic model but only gives it a new dimension. In a critique of Klein's work, York (1971) claimed that she exchanged the quantitative model of libido for the quantitative balance between the life and death drives. I would suggest that although she was not interested in the quantitative aspect of the economic model, she drew attention to the issue of balance between the drives in the dualistic model. I fully agree with O'Shaughnessy (1981) that "Klein described something similar to the pleasure principle but from another perspective... This is an object-relations perspective on the discharge of displeasurable tensions and stimuli." In my view, the above description of the "content" of the drives justifies thinking about drives from an object-relation perspective, or even more - the understanding of stable patterns of object-relation demands, situating them in the frame of theory of drives. However, it is widely understood that most Kleinians

are not interested in the biological source of the drives; what is understood by many psychoanalysts is that the drives in Kleinian metapsychology are metaphors. I believe this is over interpretation and that the Kleinian approach does not exclude the biological source of the drives.

Furthermore, biology is abandoned in the approach represented by the Lacanian tradition. For Lacan (Lacan, 1977a, pp. 235–243), the distinction between instincts and drives is crucial and the drives are completely removed from biology. In his complex elaboration, he distinguishes between need, desire, demand, and drive (Lacan, 1977a, p. 162). Drives in his theorization are purely mental entities; thus, the energy of drives seems to be truly metaphorical. Noteworthy also is his understanding of the dualism of the drives; Lacan (1988) does not differentiate death and life drives but sees destructiveness in the excessiveness and repetitiveness of drives. It accepts the economic dimension, but it does not seem possible to link this model with the original idea of drives as emerging on the “somato-psychic frontier.”

A different and interesting proposition is presented by Green (2015b). He adheres to Freud's original idea that drives have to be considered in a three-dimensional context: economic, dynamic, and topographical. Simplifying, it leads him to an understanding of the drives as forces of binding and unbinding; drives are the “powerhouse” of the development of representations linked with the quota of affect; or in other words, drives make a “link between psychic activity and the body” (Green, 1997). Penot (2017), who shares similar views, observes: “I would say that no drive trajectory can be accomplished with binding libido alone, and without benefiting from the *dissociative* force necessary for the dynamic of *organization–disorganization* that underlies the dynamic destiny of every drive pair.” I cannot do justice to this complex theorization but only want to underline a few important points for this paper. Firstly, Green suggests that drive energy is bound into representations and quotas of affects that constitute cathected objects (including a relation to an object). Secondly, such objects are not only imaginary or real persons, but also ideas. Thirdly, he sees the energy of the death drive as a force of dis-objectification/unbinding. Fourthly, he sees these two forces as oscillating throughout life (Green, 1997). Again similarly, Penot (2017) claims that without unbinding/de-cathexis, subjective life would not be possible. His point is that the development of the psyche needs both drives to balance the forces of binding and unbinding, which provides a balance between the stability of cathexis and flexibility, i.e., providing room for change. One can say there is some similarity between this approach and what was said above about Klein and her thinking about a balance between the life and death drives.

A very different view of the economic model is presented by some members of the contemporary Freudian tradition, although many contemporary Freudians do not fully share this presented view. Regarding the role and source of the drives,

Sandler and Schachter (2014) claim: “These early experiences are essentially bodily: ‘The ego is first and foremost a bodily ego’ (Freud, 1923), derived from multiple physical contacts with the mother/caretaker, and in response to the bodily pressures of the drives (hunger, thirst, cold, pain, etc.).” In this interpretation, the drives are biological impulses (like hunger), and their energy is later represented by affect. This development is described as follows: “The role of the body in its physical reality and its representation within the mind is fundamental in understanding the earliest affective experiences. The body is the seat of the drives and therefore the site of the earliest affective experiences, which over time become represented at a psychic level as the ego, and its functions develop within the matrix of the relationship with the mother” (Sandler and Schachter, 2014). In this approach, in contrast to Lacan and Klein, the emphasis is on biology and affect but drives become equated with biological needs, which, in my view, radically changes the notion of the nature and the role of drives. Although I agree that bodily impulses such as hunger, cold, and particularly impulses arising in erotogenic zones, are included in representations of drives through cathexis and constitute a very important part of these representations, I argue for the model in which the drives (in their economic sense) are forces behind cathexis, and that they are very different entities from needs. In fact, without this distinction we would not need the separate concept of the drives.

Some middle ground in this dilemma is presented by the Paris School of Psychosomatics. Aisenstein and Smadja (2010) describe their approach as follows: “In the theoretical model of the Paris school, instinctual drives have their source in bodily excitation. Their role is to deal with the tension thus created. If the sum of excitations continues to be excessive, the functional systems become disorganized and the mental apparatus overloaded.” This theoretical framework uses economic terms, considering different ways of discharging excitations and the different obstacles encountered on the way. In Green's (2010) discussion of the Paris school, he suggests that their theory of drives is too biological and raises a question about the dualistic nature of the drives. However, it is possible to interpret this approach differently; in fact, in a way that is similar to Green, by taking excitation as an energy that is different from bodily needs, leaving the nature of the energy (or excitation) unexplained.

Interestingly, Aisenstein and Smadja (2010) tried to address criticism of the abandonment of the dualism of the drives by introducing Ameisen's discovery of apoptosis as a proof of the death drive and incorporating it into their theorizing on the drives. Leaving apart the fact that apoptosis is a regulatory mechanism within a body and not a self-destructive activity, their claim makes Green's (2010) suggestion of over-biologisation somehow more valid.

However, the original Fain and Marty's (founders of the Paris School of Psychosomatics) model is explained by Fain (2018) as follows: the life drive is excitation within the body

and he believes that the nature of the death drive lies in passivity; it is fending off impulses and affect. He sees this pair of opposite drives as an important regulatory mechanism and proposed different names for them: Eros and Anteros (Fain and Braunschweig, 1971). This is a similar approach to one presented by Penot (2017) in which the opposition of drives serves a balanced development. Marty (1976) describes the development of what he calls organization (of drives) as follows: “each level of organization, the new functional ensembles, include a certain number of pre-existing, as it were constitutive functions, at the same time as a new evolving ensemble gives the functions that constitute it a new form of life;... Thus, the new evolving ensemble seems to only leave in place a kind of management which takes on a hierarchical role of great importance.” It seems to explain how representations of drives evolve and suggests their multi-layered hierarchical structure, but again leaves the question of the forces which cause those changes (the energy of the drives) unanswered. The location of this model in terms of a constantly hierarchically rebuilding organization of drives is similar to that of the Kleinian model, but it explicitly includes the economic dimension of the drives.

Further along the way toward a biologisation of the drives, we can encounter the work of Solms (2018), Solms and Panksepp (2012), Panksepp (1992). They see affects as providers of energy for developing the mind and bodily, emotional, and sensory needs, which correspond roughly to the current usage of the terms “drive,” “instinct,” and “reflex.” (Solms, 2018). This effectively equates needs felt as affects with drives, and erases the distinction between instincts, needs and drives. Although undeniably affects bear energy which constitute pressure, that energy cannot explain how it is linked with aim and object in relatively stable entity. Further, it led Solms (2021) to claim that the “drives are conscious and are in fact the source of all consciousness.” This brings an interesting but controversial spatial dimension to the theory of drives, and also makes clear sense in regard to economy. However, as we can see in Solms’ above claim and in his idea of the “Conscious Id” (Solms, 2013), he effectively departs from the model described by Freud in which drives are separate entities from needs and affects, and are never conscious (see quotations above). In his model drives have a radically different meaning, so I would prefer to consider it to be an alternative proposition to Freud’s model rather than revision of his (Freud’s) model. If we attempted to locate Solms’ model in Freud’s, we would say that the concept (traditional) of drives is unnecessary, and could be replaced with the quota of affect but again it doesn’t explain in what way affect energy is attached to specific aim and object of drives. On the other hand, if we introduced drives as a separate entity to this model, we would see Solms’ “Conscious Id” rather as an affective part of the Freudian bodily ego. In my view, this approach develops important aspects of the work of the mind, drawing on and sometimes necessarily modifying psychoanalytic knowledge; for example it would add to Contemporary Freudian (Sandler and Schachter, 2014) list

of needs relational needs. However, as regards the theory of the drives, they abandon the key assumption that drives are energies on the frontiers of soma and psyche, distinct from needs and affect; in other words, drives provide energy (real, not metaphorical) to the mental apparatus, but impulses (both external and internal), including needs, are dealt with by the mental apparatus and gradually become its content. Solms’ idea seems to be an alternative to contemporary psychoanalysis rather than an alteration which could be included in it.

In conclusion to the above overview, I am inclined toward those models which use the notion of drives as distinct from needs and affects, but also leave room for the economic dimension. From the clinical experience perspective, these models represent well what we see in consulting rooms: a variety of configurations of drives [It is also Freud’s view: “No objection can be made to anyone’s employing the concept of a drive of play or of destruction or of gregariousness when the subject-matter demands it” (Freud, 1915b)], plasticity of drives observed in analytic process and normal development alike. Equally, the economic dimension seems to me to be necessary to explain the forces of cathexis and decathexis; this is also clinically important for understanding common clinical facts like “compulsion to repeat,” transference or the process of “working through.” An additional argument against a unified list of drives is Fenichel (1935), warning against a simplistic take on the death drive (I believe it can be seen in the light of all notions of purposeful unchangeable types of drives), which can replace the analytic attitude of exploration with deterministic thinking.

In my proposition, I will use mostly the Kleinian model of unconscious phantasies and the Green model of drives. The latter is theoretically close to the Paris School of Psychosomatics and one can see also parallels in my proposition to this model. In some elements, we can also find compatibility with Lacan’s work.

Models which can be compatible with the economic model

Here, I look closer at those approaches which distinguish drives from needs and affects, and can support the model I argue for. I present only elements of these complex theories, aware that I am omitting many crucial parts of these theories and the differences between them. However, these selected elements are important for this paper’s objective as they help to address a gap in the economic aspect of the drives in the presented model.

I shall start with the Kleinian notion of unconscious phantasy from the best-known definition of unconscious phantasy: “the ‘mental expression’ of instinct (*drive in the context of this paper*¹) is unconscious phantasy. Phantasy is in

1 As already stated, in English translations of the Freudian “Trieb,” sometimes the word “instinct” is used; and sometimes “drive.” This causes confusion; I use the word “drive” to clearly distinguish it from instinct, which in other translations has a double meaning.

the first instance the mental collar, the psychic representative of instinct (*drive*). There is no impulse, no instinctual urge or response, which is not experienced as unconscious phantasy” (Isaacs, 1948). Later, Isaacs (ibidem) writes: “In early life, there is indeed a wealth of unconscious phantasies which take specific form in conjunction with the cathexis of particular bodily zones.” Isaacs also says that unconscious phantasy is “the primary content of unconscious mental processes”. My understanding of these claims is as follows: drives (as energy) are forces behind a building up of the content of the Unconscious, incorporating all what is experienced by a human, starting from bodily experiences (including affect which is perceived as bodily experience). Later, it includes all impulses, defenses, perceptions, and activities (Segal, 1964, 1994; Spillius et al., 2011b). Blass (2017) claims: “phantasies aren’t conceived of as [only] the content of the mind but rather as the material itself of which the mind is made up of.” This can be interpreted as the mind is built of elements which each contain a particular representation of momentary experience (this includes self, object, stimuli felt at the moment, need, way of satisfying need, affect etc.) perception of subsequent moments is possible only through lenses of existing phantasies; in effect, new experiences are building up newer phantasies “above” prior ones, but prior ones are part of the building material. This creates a kind of hierarchical structure where the earliest phantasies influence later ones, but new phantasies modify this entire structure by implementing new elements (of course, in some pathological states we can see that such modifications don’t take place and the individual affected is living in early primitive phantasies). In this interpretation, the unconscious phantasies are organized representations (in the Freudian sense) and affects; the latter can be understood as the vehicle for energy. However, most Kleinian writers do not pay attention to the energy dimension, but it is an important question about what forces attach different elements of experience in the organized entity of phantasy (pattern, representation) and what dynamic is involved, according to the fact that phantasies may be or may not be modified. In my view, it is a question of the nature of cathexis and decathexis. Some post-Kleinians theorize on that question, assuming that the drives have an innate purpose. I cannot find supporting evidence for this, except that primal needs and biological stimuli are obviously parts of any phantasy. However, observable and experienced purpose seem to be acquired in the process of development.

Bronstein (2015) gave her account of unconscious phantasies, elaborating on the early stages of constructing phantasies. She suggests that the most rudimentary experiences like sound, rhythm, or indeed interoceptive stimuli, are included in forming phantasy; thus, they become meaningful even before they constitute representations (in Freudian sense). My understanding of what “meaningful” means in this context would be that those experiences, through a process of cathexis, become linked with affects. Thus, they are affectively meaningful

or, in other words, underpinned with energy through cathected affect. As Solms and Panksepp (2012) have shown, needs and affects are purposeful; we can speculate that that might be the mechanism through which drives as energy acquire purpose in early stages of their development.

Bronstein (ibidem) also argues that “even though early unconscious phantasies might be modified, organized and structured in a symbolic way to form the latent content of what will become manifested as part of a coherent integrated discourse, the form which they adopt is often imbued with raw emotional components, a contribution from the paranoid-schizoid position and from primitive phantasies.” I would suggest that they are always to some extent present even if this cannot always be observed. I can support my view by Britton’s (1998a,b), Vermote’s (2014) work on Bion’s (1965) notion of transformation. They both, in different ways, argue that the development of the content of the Unconscious is a dynamic movement up and down, creating multi-dimensional space where earlier and later introjected elements influence each other. My understanding of Bronstein’s contribution is that phantasies (as the structure) can be modified by new incoming experiences but have a multi-layered structure, from primitive to well-organized symbolic substructures. They are organized hierarchically, the latest and most advanced having the biggest impact on conscious mental life. However, these layers are in constant interaction and influence each other in two ways: top-down and bottom-up. Here, I see similarities to the Paris School of Psychosomatics’ notion on the organization and development of drives described above (Marty, 1976).

Segal (1994) sees phantasy as potentially destructive and constructive at one and the same time. She believes that phantasy has the capacity to impose on reality, but also that phantasies are hypotheses for testing reality. In the context of phantasies as representations of drives, she sees duality of drives as forces of conservatism vs. adaptation. In my view, the former may be seen as a stabilizing factor of cathexis but also as an important factor in repetition compulsion (constructive and destructive respectively). The latter may be seen as a capacity to predict and a factor in the development of thinking, but a potential cause of instability, confusion and chaos if it prevails over stable cathexis (again constructive or destructive force).

This notion of unconscious phantasies provides us with a description of the content of the Unconscious and the mechanisms by which this content is built; it also describes the roles and functions that phantasy plays in mental life. Also, it can be seen as another perspective on the place of drives in the conflict between the pleasure principle (immediacy of satisfaction without a need for adaptation) and the reality principle (search for new ways for achieving satisfaction). I find here similarities between Segal’s notion on destructive vs constructive forces and Lacan’s (1977b, 1988) understanding of the Life/Death drive dilemma; both seem to see it as rather a

function borne from the intensity of the forces behind drives in the moment, than the purposeful aim included in its content.

It is worth noting that this notion of the relationship between the content of the Unconscious and conscious processes is in line with what Hopkins (2016) describes as virtual reality and the mediating role of conscious processes. My understanding of his notion is that a virtual reality is a space where phantasy/hypothesis and actual realm are negotiated [compare with Segal's (1952) writing on function of dream and imagination].

However, the Kleinian model doesn't answer our important question: what energy is propelling this process? Like Freud (1909), Klein (1927) answered these questions similarly: these processes and primal phantasies are innate and phylogenetically inherited. Followers like Money-Kyrle (1956), Bion (1962), who used the word "pre-conception" to describe it, developed ideas about the content of the Unconscious at the beginning of human life, but could not locate its source and stayed with "innate." It is very likely that a human is born with some proto-phantasies, but this doesn't explain further development, and questions about forces propelling this process remained unanswered.

Seeking for a contemporary model of drives which is compatible with free energy notion also, French psychoanalytic tradition offers a very interesting take on the drives, which seems to be the most in line with Freud's original formulation. I will use here mostly A. Green's body of work. Although André Green's (2010, 2015a) theory of the drives is significantly different from the Kleinian interpretation in terms of the structure of "content" of the drives, I do not find them contradictory.

In his theorization on the Unconscious, Green saw the representations of self, objects, and ideas as separate entities from affects. However, he claimed that they interact, and that affect plays a role as a vehicle for energy; I assume here that affect energy constitute pressure but drive energy link aim, object and affect. He emphasizes the dynamics of this structure with movements between different layers of the mind. I find particularly important his understanding of the drives as forces of binding and unbinding (investment and disinvestment, cathexis/decathexis). This could be seen as a universal model of the economic and dynamic dimensions of the fate of drives which, in my view, can be applied to the Kleinian model of unconscious phantasy as representatives of drives and their cathexes, which could be understood as a description of the content of the Unconscious. Here, it is important to notice nuanced differences in the use of the terms: representative, representations, in both bodies of work. Green (2015b) writes about three meanings of it, see below; Kleinians have a less refined definition of this term, compared with the above use of it by Issacs and Bronstein. Simplifying, Green's theory of the transformation of the drives can be described as follows:

First movement: "endo-somatic" source of energy (unknown) $\Leftrightarrow$ somato-psychic frontier (drives as energy) = representative of drives.

Second movement: Drives (somato-psychic frontier) $\Leftrightarrow$ Unconscious = representations (ideational) + quota of affects.

Third movement: Unconscious $\Leftrightarrow$ Pre-conscious = representation (word) + affects.

Fourth movement: Pre-conscious $\Leftrightarrow$ Reality = perception + action.

If we assume that the ideational representation contains a salient object and aim, we can find this model very similar to Nietzsche's description of drive (described above).

All elements of this system can and do influence each other in top-down and bottom-up directions, and also interact within the same "layer." For example, action has an impact on perception and vice versa; similarly, affect and representation. Also, perception has an impact on representation, and vice versa. This includes all directions of influence. Cathexes are guarantors of the stability of the structure. Changes to the structure demand de-cathexis. Forces behind this dynamic structuring are the drives: the life drive as a force of binding and the death drive as unbinding. Green claims that binding applies not only to object-things but also to ideational representatives. Altogether, Green's model builds a complex, multi-dimensional structure of the mind. Penot (2017), who also represents the French tradition, adds to that an important observation. He claims that only the coexistence of two drives (the forces of cathecting and de-cathecting) can promote development. Without de-cathexis, there will be stagnation and no development will be possible. Without cathexis, there will be chaos.

In terms of the dynamics of the Unconscious and a hierarchical structure of the mind, I draw the conclusion here that – in spite of differences – we can see both in the Kleinian model (particularly in Segal's and Bronstein's development) and in Green's model, a potential for compatibility with the economic model if we add to them the free energy model.

In that description, we can see the same process formulated by Green when he described the dynamics of the Unconscious and by Kleinians when they described the dynamics of Unconscious Phantasy. Therefore, I would argue that the Fristonian "hypothesis" can be seen as compatible both with Klein's notion of unconscious phantasies and with Green's understanding of the dynamics of the Unconscious.

Specifically: in Kleinian language, the "Fristonian hypothesis" describes Unconscious Phantasies including elements of constructing early (and even primal) unconscious phantasies, whereas in the Green model it would be equivalent to the representations cathected with affect, including the earliest ideational representations and the quota of affect.

What neuroscience can offer in this respect: Friston's free energy

Why did Freud fail in his attempt to find a source of free energy which gives impetus to drives and indeed to mental life? What was he seeking? His concept of drives was based on energy which demands work of the mind: "a drive appears to

us as a concept on the frontier between the mental and the somatic, as the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demand upon the mind for work" (Freud, 1915b, pp. 121–122). The crucial phrase here is *demand for work*; it could not be explained in Helmholtzian terms of free energy (which Freud tried to use) because Helmholtz's formulation described energy which is available, but Freud needed a formula for energy which demands.

Only relatively recently, the work of G. Hinton and others has allowed the development of what we can call Friston (2010) free energy. This kind of "information energy" responds well to Freud's hypothesis. It is energy which demands (a need for minimizing it), and it is energy which crosses the border between somatic and mental. That demand can be seen as a force for organizing a structure (anti-entropy) which is necessary feature of a living organism. It seems natural to seek an answer to the question: "what is the source of the drives' energy?," in the notion of free energy and predictive coding.

In recent years, the "Bayesian model of the brain" (Friston, 2008) which describes the brain as predictive agent and the "free energy principle" (Friston, 2010) have become very popular in theorizing on the interface between the brain and different aspects of the mind. These models are complex biophysical models described in mathematical language and it is not my goal here to provide an in-depth explanation. For the purpose of this paper, I need to restrict my explanation to underlining particular dimensions of these theories which I found relevant to the contemporary psychoanalytic theory of the drives. I also use the hypothesis presented by Fotopoulou and Tsakiris (2017) on the development of "the self," in which they use the "free energy principle" as the first and very convincing example of how it can be understood as a model of the somato-psychic frontier. I will claim that their work, based mainly on empirical research, describes from a new perspective phenomena which were observed by psychoanalysts.

The Bayesian model, which can be extended beyond the human brain and may be used to describe a key feature of any living entity with the ability to react to the environment (Friston, 2013), assumes that such an entity creates a hypothesis about the environment prior to perception. Such an ability is vital for survival and further development because it allows a living organism to be prepared for a changing and sometimes hostile environment. On the most basic level, an organism can react accordingly to sensory input when it needs to (e.g., by moving away from a location if the temperature is too high and for that needs prior hypothesis of such situation). Of course, it is not possible that a hypothesis generated prior to perception would always match reality completely. What happens in such a situation? One possible reaction is an action to change the environment, whereas another is to modify the hypothesis. Friston (2010) explains the mechanism of this process using the Helmholtzian theory of free energy. "Free energy" in this model

is a measure of surprise; in simple words, it is a measure of how much the hypothesis does not match reality. Free energy must be bound and can be bound either by modification of the hypothesis or by "action." Both of these reduce the gap between hypothesis and reality. It is a property of this energy which creates demands on a living organism. At a biological level, it is a mechanism which maintains homeostasis.

I placed action in quotes because it may be actual action which leads to changes in reality, but also in more complex situations contains ambiguity and different possibilities; illusion psychologically speaking, may also be an action of the mind (Brown and Friston, 2012). I have chosen not to distinguish actual action and illusion here because they both minimize free energy, so for the free energy principle they both serve the same function. However, in my proposition of the drive theory they can serve as opposite forces.

Fotopoulou and Tsakiris (2017) adopted this model in an attempt to explain one of the earliest stages in human development. They formulated a hypothesis of the earliest development based on free energy and the predictive coding model. In the process which they call "mentalization²", the structure of a minimal self is developed. They write: "In development, as ongoing intersubjective bodily interactions with the caregiver get more complex, children build increasingly more sophisticated models of their own interoceptive states, as well as strategies for minimizing free energy in the interoceptive systems." and "we have also argued that interoception, and in particular, the mentalization of interoceptive signals play a critical role in self-other boundaries. The distinction between self and other, which is crucial for self-awareness, is equally essential for awareness of other people;" finally, they conclude: "Specifically, we have described as "embodied mentalization," the process of building generative models by detecting regularities and irregularities in modality-specific and "amodal" properties, such as synchrony, and organizing sensory input of both personal and interpersonal origins into distinct, unitary multimodal schemata (perceptual inference). We have also stressed that such models refer not only to exteroception but also to interoception: the senses that inform the organism regarding the homeostatic state of the body. Furthermore, the mentalization of the body involves not only perceptual integration and subsequent inferences, but also action and thus sensorimotor integration (active inference). Accordingly, we have made the radical claim that in early infancy when the motor system is immature, proximal interactions are necessary for the active mentalization of interoceptive states and therefore the corresponding core aspects of the minimal self. There is therefore a continuity between the minimal and the interactive, social self."

2 By mentalization they understand "progressive process of build-up of multisensory predictions in perceptual inference".

They claim that such a structure contains: (1) a representation of the infant's own body including interoceptive, proprioceptive, and homeostatic signals; (2) a representation of a caring object; and (3) affect.

Now, I collate their account with Isaacs's (1948) description of the process of forming earliest phantasies: "through a natural unity of rhythm between mother and child, or through the skillful handling of any difficulties that may arise, the infant is soon able to receive pleasurable satisfaction at the breast, good co-ordination of sucking; and a positive attitude to the sucking process is set up which goes on automatically thereafter, and fosters life and health. Changes of contact and temperature, the inrush of sound and light stimulation, etc. are manifestly felt as painful. The inner stimuli of hunger, and desire for contact with the mother's body are painful, too. But sensations of warmth, the desired contact, satisfaction in sucking, freedom from outer stimuli, etc. soon bring about the actual experience of pleasurable sensation. At first, the whole weight of wish and phantasy is born by sensation and affect."

I argue here that unconscious phantasies can be translated to Friston's "prior hypotheses" (inference in Fotopoulou and Tsakiris' paper), and a situation in which any hypothesis differs from the experience is producing free energy which has to be bound. It can potentially be seen as a source of energy which is needed to build or to modify the content of unconscious phantasies. Fotopoulou and Tsakiris' description of the development of the minimal self can be seen as development of primal unconscious phantasies.

My somato-psychic frontier hypothesis

In my hypothesis presented here, I assume that the qualitative aspect of the drive in the previously presented psychoanalytic schools and Freud's original take, even if it differs in formulations, is described in compatible forms. There is an agreement that the drives' aims and objects vary and change over time in the developmental process. The drives are unconscious; conscious or unconscious [repressed] affect give them impetus, and they invisibly steer human perceptions and actions toward human satisfaction.

I claim here that whether in the form of unconscious phantasy, or a set of representations, they are equivalent to "prior hypotheses" in Fristonian free energy theory.

If Fotopoulou and Tsakiris (2017) are correct, I would extend their claim that the same mechanism should apply to all inputs; thus, all mental phenomena and experienced inputs can potentially become embodied/catheted; and affects (which are always part of those inputs) are vehicles for psychic energy. Also, I believe that this mechanism operates throughout the entire human life.

This is in line with the Kleinian formulation of unconscious phantasies, which includes and emphasizes the significance of the relation between representations of the object and the self. In later development, this structure becomes more complex because prior cathexes influence perception, consequent actions, and subsequent inputs. This process creates the circular situation that is described by both Segal and Green. The main difference in that respect between Kleinians and Green lies in the perspective of their descriptions; object-relation, and individual inner perspective respectively. Again, what remains unknown in these formulations is what factor decides what and how something becomes cathected or not during the lifespan. As we know, not every experience can be corrective in terms of inner life.

In the Friston (2013) model, any living organism creates a hypothesis about the environment starting from very rudimentary data such as, for example, homeostasis; and any new change in a given environment releases free energy. During the earliest stages of human development, it must include mostly signals from the body: interoceptive, viscerosensitive, and homeostatic states, but also elementary external inputs from objects or rather part-objects like sounds, touch. The dominance of bodily impulses in this stage explains the fact that the earliest hypotheses and mechanisms reflect bodily states; they constitute the bodily ego. I believe this process starts in the prenatal stage and the infant is born with some rudimentary hypotheses. Those prenatal hypotheses have been seen as innate or phylogenetic phantasies. They are heavily challenged after birth by a completely changed environment (both external and internal). Since development allows this possibility, it includes a richer spectrum of signals from the affective system and the external environment. I believe that Kristeva (1984, 2000) describes the same process from a different perspective. She believes that at first an infant experiences external input (including the mother) as sound, voice, rhythm etc. She calls it semiotic function and believes that together with bodily experiences and affect, they constitute "emotional experience, both psychical and subjective, based on the drives in an interpersonal context" (ibidem, 2000). This sounds very similar to Fotopoulou and Tsakiris (2017). Bronstein (2015) uses Kristeva's work to illuminate the process of formatting the earliest unconscious phantasies and their role in mental functioning; she writes (ibidem, 2015) "In my opinion, these semiotic aspects that are part of the early relationships and phantasies of the mother's body play a very important part in how we experience the world and communicate with others. Issues such as the patient's rhythm of speech, the 'musicality' in their verbal expression, the intensity placed on their words, sensitivity to noises and to silences, and issues of space in relation to the analyst's physicality can exercise a powerful effect in the session." Both Kristeva and Bronstein believe these aspects of phantasy (thus drives) remain unconscious; Kristeva calls it the "pre-narrative envelope." I add that the content of these

phantasies (envelopes) is projected onto perception, emotional experiences, more mature aspects of mind like thinking and all other aspects of mental life.

I believe that the “envelope metaphor” applies well to all unconscious processes and phantasies; also, despite the fact that phantasies (and the whole Unconscious) operate as a whole hierarchical structure, there are also compartments within it; thus, we can say there are a number of envelopes there. The same applies to the mind as a whole, with the most important barrier being between the Conscious and the Unconscious. There are many notions of such barriers, starting from Freud’s “stimulus barrier,” Federn’s (1956) ego boundaries, and Bion’s (1962) “psychic membrane,” among many others. Rabeyron (2021) formulates an interesting hypothesis linking these psychoanalytic notions with the concept of Markov blankets, which is an important property of the Friston principle of free energy. In both psychoanalysis and Friston (2013) theory, the notion of a boundary between the outside and the inside of the system is crucial. It determines the ability to maintain an organically consistent system, but it also determines differentiation between the inside and outside environment. Furthermore, maintaining the organized system demands the ability to react to changes in the outside environment; this requires building hypotheses and reacting accordingly, but of course it also requires differentiation between the inner and outer system. Only then we do have a gap between hypothesis and reality which creates free energy.

From an analytic point of view, the most important barrier is between the inner world and external reality. The task of dealing with this is, according to Freud (1911), Ego function; and the process of adjusting to reality is called the “reality principle.” However, Freud (1911) writes:

“A general tendency of our mental apparatus, which can be traced to the economic principle of saving expenditure [of energy], seems to find expression in the tenacity with which we hold on to the sources of pleasure at our disposal, and in the difficulty with which we renounce them. With the introduction of the reality principle, one species of thought-activity was split off; it was kept free from reality-testing and remained subordinated to the pleasure principle alone. This activity is *phantasying*,”

I propose the following interpretation of this: saving energy, which for Freud is a basic principle, can be achieved in two ways. The first is through the “reality principle” which includes two elements: action to change reality or adjusting our goals and objects to reality (modification of our drives’ vicissitudes/hypothesis). The second way of saving energy can be creating an illusion [what Freud (1911) called hallucinatory wish fulfilment] that reality is exactly like our phantasy/drives’ vicissitudes/hypothesis. The latter serves the pleasure principle without the need to modify inner representations, but it needs to create a barrier which is the barrier between the Conscious and Unconscious. This barrier creates a “new

frontier;” my understanding of this new frontier is the place where “the conflict between Ego (which represent reality) and Id” (Freud, 1920, 1923) takes place. Then we have another sub-system Super-Ego developed through cathexis and identification (Freud, 1923), which demands a new barrier. All new barriers are potential for conflicts and may create free energy, in Freud’s words: tension. Therefore, the Ego can mitigate potential conflicts between different agencies of the mind but also through defense mechanisms and creating illusions can create new frontiers which generate conflicts and potentially aggravate discrepancy between hypothesis/phantasies and reality.

As the body and the environment are dynamic structures, the organism (mind and body) is exposed to a certain amount of free energy most of the time, despite all efforts to minimize it.

Furthermore, I suggest that with growing complexity the initial conflict between reality and internal hypotheses/unconscious phantasies is supplemented by conflicts between different agencies of the mind; later by split-off or dissociated parts of the mind etc. (all these new sub-systems have to produce their own hypotheses at the same time as being hypotheses themselves). All this constitutes the conflictual character of the mind.

I also suggest that this process involves defense mechanisms which in this context can be described as agents of misrepresentation or creating illusion, which is to minimize free energy. From a psychoanalytic point of view, illusion is necessary to maintain the mind intact by negotiating internal conflicts, but can also be a cause of pathology when illusion takes over the reality principle (Steiner, 2020). From an energy-economic point of view (at least in Friston’s notion of it), an illusion can minimize free energy. Brown and Friston (2012) argue that “illusory percepts are, in fact, Bayes-optimal, and represent the most likely explanation for ambiguous input.” In other words, in a complex and ambiguous situation, the mind can produce an illusory explanation which can minimize free energy but falsely match a hypothesis with reality. Later, it has the potential to become part of another layer of internal hypotheses or, in analytic terms, to be cathected. Complexity and ambiguity may be created not only from overwhelming external input but also from conflicts between external input and already cathected representations of objects or wishes, and from internal conflicts between different psychic agencies or contradictory feelings. Those illusory ideas can also be cathected and become parts of unconscious phantasies. I suggest that such a process lays the ground for the Ego function that Lacan called “méconnaissance” (Lacan, 2014). During development, this function becomes one of the possible purely psychic “actions” alongside motor actions, which become behavioral and often serve as a means of communicating wishes to change, or adapt reality to the prior hypothesis. An example of a psychoanalytic description of such a situation may be any account of projective identification which may serve as only an illusion, but can also be actualized by forcing a real object to play a role prescribed

in phantasy. A more complex account of the complexity of a relationship when a compromise between phantasy and reality of the object has to be found is in Joseph Sandler's notion of role responsiveness (Sandler, 1976). The highest functional level of this process involving illusion is what Bion (1963) described as the K and -K dimension; where K stands for thoughts which represent reality (inner or outer) and -K stands for thoughts which represent a false narrative.

I believe that my hypothesis on drive energy can successfully complement contemporary psychoanalytic theory of drives and Friston's theory of free energy in a similar way to how Fotopoulou and Tsakiris used it. In effect, it may constitute the model that adds to contemporary psychoanalytic views on drives an economic point of view containing a biological (in fact, biophysical) source of the drives' energy. In summary, I hypothesize that the free energy postulated by Friston is the primal source of that energy. Furthermore, I argue that Fotopoulou and Tsakiris' hypothesis can explain what energy, and in what way it is used to build the content of the qualitative dimension of the drives.

I postulate that the unknown "endo-somatic source of energy" may be Fristonian free energy, and the processes of cathexis and de-cathexis, as described by Green, can be seen as processes of minimizing (binding) free energy.

As mentioned above, the human mind is a very complex apparatus which contains subsystems and thus creates possible barriers/frontiers not only at the level of the biological body and environment, but also between purely psychical entities. This suggests that possible sources of free energy may have different locations. The free energy principle teaches us that free energy must be minimized by binding. This is achieved by two-way traffic. One direction is action: on a bodily level, for instance, it might be a release of hormones or any other somatic reaction; on the external level, it might be a motor reaction or illusion (including what Freud, 1900 described as hallucinatory wish fulfilment and negative hallucination). The other direction is the development of hypotheses. I prefer to say development instead of modification because, with time, a system of hypotheses becomes more and more complex. There are more data to deal with, while the ability to deal with such a growing magnitude of input is also growing. I suggest that those earliest hypotheses are what psychoanalysts call primary phantasies. They include bodily ego, primal objects, part-object, object-relations and so on. Different schools of psychoanalysis lay emphasis on different aspects of that process. I also suggest that the ability to develop those phantasies into a more mature form and to react adequately to a dynamic environment (internal and external) is the equivalent of what psychoanalysts call the Ego. The processes of modifying those internal hypotheses and actions to actualize them can be seen as the processes of de-cathexis and cathexis respectively. This happens on different levels of the mind, but the main "economic" mechanism remains the same.

Input in this process is the entire life experience; by life experience, I understand here a "product" created by perception and unconscious phantasies, which influence perception. This includes affect which is part of both perception and unconscious phantasy. Perception in this model may or may not be conscious. All this creates new hypotheses which make links between all experiences within a given moment. With maturation, those hypotheses should become more specific and selective; thus, hypotheses become more adequate. This process happens by building a hierarchical structure with more primitive and general content at the bottom, and more specific and complex content at the top. In psychoanalytic language, bottom stands for deep unconscious and top for conscious. However, this developmental process can be arrested by over-cathexis; a situation that one or more hypotheses/phantasies become dominant and rigidly defended against any modification. Such a defense can be carried out by means of illusion and/or behaviorally.

I believe this developmental process is the process that Kleinians call the development of unconscious phantasies. I suggest that the process of making hypotheses or unconscious phantasies more selective includes the creation of representations and linking them with affect and simultaneously developing defensive mechanisms. In this way, representations of objects, representation of self, ideas and defensive strategies, become cathected and charged with affect, constituting complex hypotheses about life, the world and oneself. However, those hypotheses have ingrained in them the demand to be confirmed and cannot be given up, but as a whole structure can be gradually modified.

Any change in such structures needs de-cathexis. In other words, adaptation to a changing world demands a constant balancing of cathexis and de-cathexis. I am in agreement here with both Green and Penot that de-cathexis plays a crucial part in the processes of any psychic development such as subjectivization or mourning. Their theorizing on this issue is in line with Britton's (1998c) observation that mourning and development requires a relinquishment of beliefs (ideas or objects). However, I would go further and claim that de-cathexis has a part in every and any process of restructuring unconscious phantasies. In my understanding, it means changes to representations of objects, self, and so on.

However, it confronts us with an important paradox; on the bodily (biological) level, maintaining homeostasis is the main factor keeping an organism alive, so maintaining a low level of free energy is desirable. Although, in terms of "psychic life," "Nirvana principle" (Freud, 1920), also called "inertia principle," which is a tendency to keep tension/free energy constantly at low a level is an anti-life force (death drive), free energy and dynamics cathexis/de-cathexis/new cathexis is a condition of "*sine qua non*" for any psychic development.

In this model, the energy of Freudian drives in its "pure" form is "Fristonian" free energy. As described by Friston, two

ways of binding energy make drives dual; I can call them simply forces of cathexis – de-cathexis. At the start of this process, this free energy is a biophysical phenomenon and an intrinsic feature of any living entity; also intrinsic is the necessity to minimize it in two ways simultaneously, as described above.

In the relatively long time which is needed to achieve a high level of complexity in human experience, they become not only dual but also dialectic in their character (Segal, 1993). The drives are necessary for creating the potential to develop the mind, but they can also restrain this potential.

Both sides of this dialectic pair are absolutely crucial. The dialectic character of the drives is essential because only these two forces operating together can support development; as Penot observed, the permanent dominance of either of the two can arrest development. We have many clinical accounts of pathology coming from such arrests. For example, Bell (2015) tries to classify phenomenologically some of the clinical expressions of such a situation under the aegis of the death drive. In such a context, I would suggest that what he calls the death drive is either the dominance of cathexis in the form of living in phantasy (trying to impose it on the external world or subject is isolating from the world), or dominance of decathexis in the form of a chaotic, fragmented and split mental life. Feldman's (1997) account of projective identification in the session is, in my view, captured exactly in this line: "From the patient's point of view, the projections represent an attempt to reduce the discrepancy between the phantasy of some archaic object relationship, and what the patient experiences in the analytic situation." I think it is a situation when the patient's psychic apparatus cannot modify its hypothesis and tries to impose hypothesis on reality, in the framework of my proposition it is a situation when the balance between drives is distorted toward cathexis. I also fully agree with Bell in his clinical observation and argument, however, I would suggest that on a metapsychological level what he calls death drive is a severe imbalance between two dialectical forces, rather than the "death drive" as a sole force.

I emphasize their dialectic character, proposing "balancing" instead of Green's term "oscillation" (Green, 2015a), because they operate at the same time and have the same source of free energy. When observing the vicissitudes of the drives, I believe that what we see as fusion and de-fusion are in fact different configurations and stages in balancing between the two ways of minimizing free energy, which initially cannot be dual in its nature and achieves a dual character in the process of minimizing and transforming it into the drives' energy. Another important factor in this development is that the human mind also uses "the third way" of minimizing free energy, which is illusion (which in a broad sense is the equivalent of deceiving one's own consciousness); although we have to remember that the broad church of "illusion" includes some potentially pathological mechanisms such as massive projective identifications, but also defense mechanisms which can preserve

the balance of the mind in excessively complex situations. Developmental mechanisms such as symbolization, which is crucial for mature functioning, also include elements of illusion (compare with Rabeyron, 2021) etc. To capture all these nuances will demand a separate body of work but in this paper I am proposing only the framework.

Another crucial factor is the time needed to achieve complexity in these internal hypotheses. This is due not only to the maturation of the brain but also, as clinical observations teach us, because changes in unconscious phantasies can be made only gradually. Perhaps too much surprise (free energy) is so overwhelming to the system that the balance between the two directions of the drives is disturbed to a degree that arrests development. I would suggest that this might be the biophysical equivalent of what Winnicott (1965) called impingement, and what lies behind the overexcitement described by theoreticians of the Paris School of Psychosomatics.

Possible further directions for development

If my hypothesis is correct, it might open some new possibilities for work on metapsychology. In my view, there are at least five issues, as follows:

(1) We might consider the biggest developmental crises to be underpinned by significant changes in the internal and external environment, which lead to higher levels of free energy so that, subsequently, we observe a higher pressure in the drives. For example, this could explain the fact that adolescent crises are experienced at different ages by different persons and do not always occur at the same time as biological maturation (e.g., hormone activity).

Similarly, we can think about the psychoanalytic situation exactly in the way that Scarfone (2018) writes that psychoanalysis: "presents patients with new packets of free energy."

(2) One of the controversial topics in psychoanalysis is the location of aggression in drive theory: i.e., the question of if or when aggression is self-preserving; if and when it is a derivative of the "death drive;" and so on. In my hypothesis, we can see aggression as an affective element of the unconscious phantasy/hypothesis, which could be considered to be part of a complex structure and not rigidly attributed to any of the drives. However, aggression as affect and as behavior can be linked to the "action" side of the dialectical pair of drives; thus, rather as a force of conservatism in Segal's words. Nevertheless, this does not qualify aggression as a negative force.

Destructiveness can be linked with aggression or passivity alike but not directly and not always. I would suggest that qualifying a phantasy as destructive should depend on the role it plays in the dimension of adaptation/development vs. stagnation/maladaptation, rather than on descriptive features.

(3) Another concept arousing discussion is the infant's envy, particularly in the earliest stages of human life. In the light of my proposition, envy seems to be a derivative of a way of minimizing free energy: an obvious consequence of the massive challenges of an environment which is changing fast and radically. The infant moves from the relatively stable environment of the womb to a much more changeable life. An infant is exposed to more and more new perceptions which cannot as yet have counterparts in internal hypotheses (e.g., the need to scream when a hunger is a new experience). The most crucial challenge seems to be a clash between experienced dependency and the omnipotent solitude of a baby in the womb. This kind of reaction (an envious wish to destroy) leading to the destruction of a reality that is too new seems to be inevitable. However, Winnicott's (1955) claim concerning the role of the environment should be taken into account. I agree that the degree of this challenge (and thus the amount of free energy that drives envious reaction) can be mitigated or increased from outside; namely, by a mothering person.

(4) As Bronstein (2011) noted, one of the most important differences between approaches in psychosomatics is a different understanding of the role of drives and different models of drive theory. This new proposition may help discussion in the psychosomatic field. One example is the question of whether psychosomatic symptoms are part of unconscious phantasies, or a direct product of the death drive which blocks the development of psychic structure. Another example is the question of whether psychosomatic symptoms are caused by excessive drive energy, or rather by an imbalance of drives which arrested development on the somatic experience level, and prevented symbolization.

(5) Additionally, it might be possible in further theorizing to explain, in terms of mathematical logic, Bion's (1963) concepts of containment and "grid" using Tarski's (1933) undefinability theorem and his work on language and the definition of truth, which would help us to understand predictive coding in a relational context.

Conclusion

Firstly, I need to conclude that despite the fact that we have various descriptions of a role and the dynamics of the derivatives of the drives, it is common to all psychoanalytic schools of thought that we cannot have access to the drives as such. Taking seriously the idea that all mental activities are underpinned and powered by drives, I find both Green's take on the drives and the Kleinian notion of Unconscious Phantasies as representative of the drives very convincing. It comprehends all aspects of transformations and various forms of the drives' vicissitudes. Bronstein's (2015) contribution, which I see as a multi-layered hierarchical model, makes it much more compatible with other schools' approaches, such as the contemporary Freudians' focus

on ego development (Sandler and Schachter, 2014), or those that focus on sensitivity to input from the environment (Winnicott, 1965). Segal's (1994) contribution describes a particular function of unconscious phantasies as internal hypotheses; consequently, she maps out the nature of the interface between those internal hypotheses and reality.

This paper has argued that free energy as described by Friston can be seen as the source of the drives' energy in psychoanalytic theory. Similarly, the minimizing of free energy can be seen as binding it; this process would be the Freudian "somato-psychic frontier." In this model, tension can be understood as a high level of free energy created by a gap between reality and prior hypotheses/unconscious phantasies (also between different "departments" of the mind). A good example of this would be Bion's (1979) "emotional storm," which in his view is the unavoidable tension when two personalities, previously unknown to each other, meet. If cathected, needs, impulses, instincts (in the biological sense), and ideas become parts of vicissitudes of drives; the drives acquire aims and objects in that process, and are organized in the form of unconscious phantasies/hypotheses. Subsequently, we can understand the "pleasure principle" as a principle minimizing free energy; thus, the state of pleasure as a low level of free energy and displeasure as a high level of free energy. The "Reality Principle" would be necessary to achieve a high level of adequacy between unconscious phantasies and reality, which also minimizes free energy but in a different more adaptive way. The role of erotogenic zones can be understood as sources of new and early significant impulses cathected and often linked with positive affect; in this way, they are incorporated into unconscious phantasies and achieve dominant status at a given time, but might be modified later by further decaathesis and cathexis. I see this proposition as a complement to the current psychoanalytic theory of the drives; in particular, Kleinian and Green's understanding of the drives. It might make the economic dimension of Freudian theory an important part of psychoanalytic metapsychology again.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author would like to thank my colleagues: Steven Groarke, Jim Hopkins, Cezary Zechowski, and Francois Loew for discussing my ideas with me.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aisenstein, M., and Smadja, C. (2010). Conceptual framework from the paris psychosomatic school: a clinical psychoanalytic approach to oncology. *Int. J. Psychoanal.* 91, 621–640. doi: 10.1111/j.1745-8315.2010.00256.x
- Bell, D. (2015). The death drive: Phenomenological perspectives in contemporary Kleinian theory. *Int. J. Psychoanal.* 96, 411–423. doi: 10.1111/1745-8315.12212
- Bell, D. (2017). Unconscious phantasy: Some historical and conceptual dimensions. *Int. J. Psychoanal.* 98, 785–798. doi: 10.1111/1745-8315.12586
- Bion, W. R. (1962). *Learning from Experience*. London: Heinemann.
- Bion, W. R. (1963). *Elements of Psycho-Analysis*. London: Karnac Books.
- Bion, W. R. (1965). *Transformation. Change from Learning to Growth*. London: William Heinemann Medical Books.
- Bion, W. R. (1979). "Making the best of a bad job," in *Clinical Seminars and Other Works*, Ed. F. Bion (London: Karnac Books).
- Blass, R. (2017). Reflections on Klein's radical notion of phantasy and its implications for analytic practice. *Int. J. Psychoanal.* 98, 841–859. doi: 10.1111/1745-8315.12674
- Britton, R. (1998a). Taming Wild Thoughts. By W. R. Bion. Ed. by Francesca Bion. London: Karnac Books. 1997. *Int. J. Psychoanal.* 79, 817–819.
- Britton, R. (1998b). *Belief and Psychic Reality in Belief and Imagination*. London: Routledge, 8–18.
- Britton, R. (1998c). *Before and After the Depressive Position $Ps(n) > D9n > Ps(n+1)$ in Belief and Imagination*. London: Routledge, 69–81.
- Britton, R. (2015). *Between Mind and Brain. Models of the Mind and Models of the Mind*. London: Karnac Books.
- Bronstein, C. (2011). On Psychosomatics: the search for meaning. *Int. J. Psychoanal.* 92, 173–195. doi: 10.1111/j.1745-8315.2010.00388.x
- Bronstein, C. (2015). Finding unconscious phantasy in the session: Recognizing form. *Int. J. Psychoanal.* 96, 925–944. doi: 10.1111/1745-8315.12330
- Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the Cornsweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043
- Cambridge English Dictionary (2022). Available online at: <https://dictionary.cambridge.org/dictionary/english/source> (Accessed on August 13, 2022).
- Conrad, J. A. (2021). Drive theory, redux: a history and reconsideration of the drives. *Int. J. Psychoanal.* 102, 492–518. doi: 10.1080/00207578.2020.1848389
- Fain, M. (2018). Mentalisation and passivity. *Int. J. Psychoanal.* 99, 479–484. doi: 10.1080/00207578.2017.1401434
- Fain, M., and Braunschweig, D. (1971). *Eros et Anteros*. Paris: Payot.
- Federn, P. (1956). *Ego-Psychology and the Psychoses*. New York, NY: Basic Books.
- Feldman, M. (1997). Projective Identification: the analyst's involvement. *Int. J. Psychoanal.* 78, 227–241.
- Feldman, M. (2000). Some Views on the Manifestation of the Death Instinct in Clinical Work. *Int. J. Psychoanal.* 81, 53–65. doi: 10.1516/0020757001599528
- Fenichel, O. (1935). "A critique of the death instinct," in *Collected Papers*, Ed. L. V. Ahlfors (London: Routledge), 363–372.
- Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: The social origins of interoceptive inference. *Neuropsychanalysis* 19, 3–28. doi: 10.1080/15294145.2017.1294031
- Freud, S. (1900). *The Interpretation of Dreams*, Standard Edn, Vol. IV. London: Hogarth Press.
- Freud, S. (1905). *Three Essays on the Theory of Sexuality*, Standard Edn, Vol. VII. London: Hogarth Press.
- Freud, S. (1909). *Analysis of a Phobia in a Five-Year-Old Boy*, Standard Edn, Vol. X. London: Hogarth Press, 2–149.
- Freud, S. (1911). *Formulation on the Two Principles of Mental Functioning*, Standard Edn, Vol. XII. London: Hogarth Press, 218–226.
- Freud, S. (1915a). *The Unconscious*, Standard Edn, Vol. XIV. London: Hogarth Press, 181.
- Freud, S. (1915b). *Instincts and their Vicissitudes*, Standard Edn, Vol. XIV. London: Hogarth Press, 122.
- Freud, S. (1915c). *Repression*, Standard Edn, Vol. XIV. London: Hogarth Press.
- Freud, S. (1920). *Beyond the Pleasure Principle*, Standard Edn, Vol. XVIII. London: Hogarth Press.
- Freud, S. (1923). *The Ego and the Id*, Standard Edn, Vol. XIX. London: Hogarth Press, 26.
- Freud, S. (1933). *New Introductory Lectures on Psychoanalysis*, Standard Edn, Vol. XXII. London: Hogarth Press, 106.
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211
- Friston, K. (2010). The free-energy principle: A unified brain theory. *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. (2013). Life as we it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475
- Friston, K., and Stephan, K. E. (2007). Free energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y
- Green, A. (1997). *On Private Madness*. London: Karnac classics, 174–213.
- Green, A. (2010). "Thoughts on the paris school of psychosomatics," in *Psychosomatics Today*, eds M. Aisenstein and E. Rappoport de Aisemberg (London: Karnac).
- Green, A. (2015a). "Binding and recognition," in *Key Ideas for a Contemporary Psychoanalysis*, ed. A. Green (Hove: Routledge), 185–186.
- Green, A. (2015b). "The analysis of the material into its component parts," in *Key Ideas for a Contemporary Psychoanalysis*, ed. A. Green (Hove: Routledge), 125–165.
- Hopkins, J. (2016). Free energy and virtual reality in neuroscience and psychoanalysis: a complexity theory of dreaming and mental disorder. *Front. Psychol.* 7:922. doi: 10.3389/fpsyg.2016.00922

- Isaacs, S. (1948). The nature and function of phantasy. *Int. J. Psychoanal.* 29, 73–97.
- Jones, E. (1953). *The Life and Work of Sigmund Freud*, Vol. I. London: Hogarth Press.
- Klein, M. Ed. (1927). “Criminal tendencies in normal children,” in *The Writings of Melanie Klein*, Vol. 1. (London: Hogarth Press), 170–185. doi: 10.1111/j.2044-8341.1927.tb00652.x
- Klein, M. Ed. (1940). “Mourning and its relation to manic-depressive states,” in *The Writings of Melanie Klein*, Vol. 1. (London: Hogarth Press), 244–369.
- Kristeva, J. (1984). *Revolution in Poetic Language*. New York, NY: Columbia University Press.
- Kristeva, J. (2000). From symbols to flesh: the polymorphous destiny of narration. *Int. J. Psychoanal.* 81, 771–787. doi: 10.1516/0020757001600147
- Lacan, J. (1977a). “The seminars book XI,” in *The Four Fundamental Concepts of Psychoanalysis 1964*, ed. A. Sheridan (London: Hogarth Press), 235–243. Trans.
- Lacan, J. (1977b). *Écrits: A Selection*, ed. A. Sheridan (London: Tavistock Publications), 301. Trans.
- Lacan, J. (1988). *The Seminars Book I: Freud's Papers on Technique 1953–1954*, ed. J. Forrester (Cambridge: Cambridge University Press), 118–120. Trans.
- Lacan, J. (2014). *The Seminars Book X: Anxiety ('L'Angoisse' 1962–63)*. Cambridge: Polity Press.
- Laplanche, J., and Pontalis, J.-B. (1973a). “Economic,” in *The Language of Psychoanalysis. Reprinted 2006*, Trans. D. N. Smith (London: Karnac Books), 127.
- Laplanche, J., and Pontalis, J.-B. (1973b). “Instinct (or Drive),” in *The Language of Psychoanalysis. Reprinted 2006*, Trans. D. N. Smith (London: Karnac Books), 214.
- Marty, P. (1976). *Les Mouvements Individuels de vie et de Mort*. Paris: Payot.
- Money-Kyrle, R. (1956). “The world of the unconscious and the world of common sense,” in *The Collected Papers of Roger Money-Kyrle. Reprinted 1978*, eds D. Meltzer and E. O'Shaughnessy (Strath Tay: Clunie Press), 318–329.
- Nietzsche, F. (1881). *2003 Daybreak*, eds M. Clark and B. Leiter (Cambridge: Cambridge University Press).
- O'Shaughnessy, E. (1981). A commemorative essay on W. R. Bion's theory of thinking. *J. Child Psychother.* 7, 181–192. doi: 10.1080/00754178108255031
- Panksepp, J. (1992). A critical role for affective neuroscience in resolving what is basic about basic emotions. *Psychol. Rev.* 99, 554–560. doi: 10.1037/0033-295X.99.3.554
- Penot, B. (2017). The So-Called Death Drive, an Indispensable Force for Any Subjective Life. *Int. J. Psychoanal.* 98, 299–321. doi: 10.1111/1745-8315.12610
- Rabeyron, T. (2021). Beyond the death drive: Entropy and free energy. *Int. J. Psychoanal.* 102, 878–905. doi: 10.1080/00207578.2021.1932514
- Sandler, A.-M., and Schachter, J. (2014). An outline of our contemporary freudian approach. *Brit. J. Psychother.* 30, 6–17. doi: 10.1111/bjp.12059
- Sandler, J. (1976). Countertransference and role-responsiveness. *Int. R. Psychoanal.* 3, 43–47.
- Scarfone, D. (2018). Free association, surprise, trauma and transference. *Psychoanal. Inq.* 38, 468–477. doi: 10.1080/07351690.2018.1480232
- Schneider, G. H. (1880). *Der Thierische Wille*. Leipzig: A. Abel.
- Schopenhauer, A. (1958). *The World as Will and Representation*, Trans. E. F. Payne (New York, NY: Dover Publications).
- Segal, H. (1952). *The Nature and Function of Phantasy in Development in Psychoanalysis*, eds M. Klein, P. Heimann, S. Isaacs, and J. Riviere (London: Hogarth Press).
- Segal, H. (1964). Symposium on phantasy: Fantasy and other mental processes. *Int. J. Psychonal.* 45, 191–194.
- Segal, H. (1985). “Chapter3: The Klein-Bion model,” in *Models of the Mind: Their Relationships to Clinical Work*, Vol. 81, ed. A. Rothstein (New York, NY: International University Press), 35–47.
- Segal, H. (1993). On the clinical usefulness of the concept of the death instinct. *Int. J. Psychoanal.* 74, 55–61.
- Segal, H. (1994). Phantasy and reality. *Int. J. Psychoanal.* 75, 395–401.
- Solms, M. (2013). The Conscious Id. *Neuropsychanalysis* 15, 5–19. doi: 10.1080/15294145.2013.10773711
- Solms, M. (2018). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9:2714. doi: 10.3389/fpsyg.2018.02714
- Solms, M. (2021). Revision of drive theory. *J. Am. Psychoanal. Ass.* 69, 1033–1091. doi: 10.1177/00030651211057041
- Solms, M., and Panksepp, J. (2012). The ‘Id’ knows more than ‘Ego’ admits: Neuropsychanalytic and primal consciousness perspectives on the interface of affective and cognitive neuroscience. *Brain Sci.* 2, 147–175. doi: 10.3390/brainsci2020147
- Spencer, H. (1870). *The Principles of Psychology*. London: Williams and Norgate. doi: 10.1037/12297-000
- Spillius, E. B., Milton, J., Garvey, P., Couve, C., and Steiner, D. eds (2011a). “Economic model,” in *The New Dictionary of Kleinian Thought*. (London: Routledge), 316–319.
- Spillius, E. B., Milton, J., Garvey, P., Couve, C., and Steiner, D. eds (2011b). “Unconscious phantasy,” in *The New Dictionary of Kleinian Thought*. (London: Routledge), 3–15.
- Steiner, J. (2020). *Illusion, Disillusion, and Irony in Psychoanalysis*. London: Routledge. doi: 10.4324/9781003030508
- Strachey, J. (1957). *Instincts and their Vicissitudes –Editor's Note*, Standart Edn, Vol. XIV, ed. S. Freud (London: Hogarth Press), 111–116.
- Tarski, A. (1933). *Pojecie Prawdy w Językach Nauk Dedukcyjnych*. Warszawa: Towarzystwo Naukowe Warszawskie.
- Vermote, R. (2014). Transformations et transmissions du fonctionnement psychique: approche integrative et implications cliniques. *Rev. Française Psychanal.* 78, 389–405. doi: 10.3917/rfp.782.0389
- Winnicott, D. W. (1955). Metapsychological and clinical aspects of regression within the psychoanalytical set-up. *Int. J. Psychoanal.* 36, 278–294.
- Winnicott, D. W. Ed. (1965). “Ego integration in child development,” in *The Maturation Process and the Facilitating Environment*. (London: Hogarth Press), 56–60. doi: 10.4324/9780429482410-4
- York, C. (1971). Some suggestions for a critique of Kleinian psychology. *Psychoanal. Study Child* 26, 129–155. doi: 10.1080/00797308.1971.11822268



OPEN ACCESS

EDITED BY

Filippo Cieri,
Neurological Institute, United States

REVIEWED BY

Andrea Scalabrini,
University of Bergamo, Italy
Elena V. Mnatsakanian,
Institute of Higher Nervous Activity
and Neurophysiology (RAS), Russia

*CORRESPONDENCE

Cristina Trentini
cristina.trentini@uniroma1.it

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 05 September 2022

ACCEPTED 02 November 2022

PUBLISHED 28 November 2022

CITATION

Tanzilli A, Trentini C, Grecucci A,
Carone N, Ciacchella C, Lai C,
Sabogal-Rueda MD and Lingardi V
(2022) Therapist reactions to patient
personality: A pilot study of clinicians'
emotional and neural responses using
three clinical vignettes from
in treatment series.
Front. Hum. Neurosci. 16:1037486.
doi: 10.3389/fnhum.2022.1037486

COPYRIGHT

© 2022 Tanzilli, Trentini, Grecucci,
Carone, Ciacchella, Lai,
Sabogal-Rueda and Lingardi. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Therapist reactions to patient personality: A pilot study of clinicians' emotional and neural responses using three clinical vignettes from in treatment series

Annalisa Tanzilli¹, Cristina Trentini^{1*}, Alessandro Grecucci²,
Nicola Carone³, Chiara Ciacchella¹, Carlo Lai¹,
Miguel David Sabogal-Rueda⁴ and Vittorio Lingardi¹

¹Department of Dynamic, Clinical Psychology, and Health Studies, Faculty of Medicine and Psychology, Sapienza University of Rome, Rome, Italy, ²Clinical and Affective Neuroscience Lab, Department of Psychology and Cognitive Sciences, University of Trento, Rovereto, Italy, ³Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy, ⁴Department of Psychology, Catholic University of Sacred Heart, Milan, Italy

Introduction: Therapists' responses to patients play a crucial role in psychotherapy and are considered a key component of the patient–clinician relationship, which promotes successful treatment outcomes. To date, no empirical research has ever investigated therapist response patterns to patients with different personality disorders from a neuroscience perspective.

Methods: In the present study, psychodynamic therapists ($N = 14$) were asked to complete a battery of instruments (including the Therapist Response Questionnaire) after watching three videos showing clinical interactions between a therapist and three patients with narcissistic, histrionic/borderline, and depressive personality disorders, respectively. Subsequently, participants' high-density electroencephalography (hdEEG) was recorded as they passively viewed pictures of the patients' faces, which were selected from the still images of the previously shown videos. Supervised machine learning (ML) was used to evaluate whether: (1) therapists' responses predicted which patient they observed during the EEG task and whether specific clinician reactions were involved in distinguishing between patients with different personality disorders (using pairwise comparisons); and (2) therapists' event-related potentials (ERPs) predicted which patient they observed during the laboratory experiment and whether distinct ERP components allowed this forecast.

Results: The results indicated that therapists showed distinct patterns of criticized/devalued and sexualized reactions to visual depictions of patients with different personality disorders, at statistically systematic and clinically meaningful levels. Moreover, therapists' late positive potentials (LPPs) in the hippocampus were able to determine which patient they observed during the EEG task, with high accuracy.

Discussion: These results, albeit preliminary, shed light on the role played by therapists' memory processes in psychotherapy. Clinical and neuroscience implications of the empirical investigation of therapist responses are discussed.

KEYWORDS

therapist reaction, patient personality, neural response, ERPs, memory, hippocampus, LPPs, psychotherapy

Introduction

The therapeutic relationship represents one of the most important mutative factors of good treatment outcome (Wampold, 2015; Norcross and Lambert, 2019). Of note, therapists' emotional, cognitive, motivational, and behavioral response patterns to patients (in this context, we use the terms "therapist's response" and "therapist's reaction" interchangeably) play a crucial role in psychotherapy. Furthermore, they are a critical component of the patient–therapist relationship, which is strongly related to multifaceted processes involved in patient change (Gabbard, 1995; Gabbard and Westen, 2003; Hayes et al., 2018).

The theoretical-clinical roots of this relational dimension can be traced to the classical psychoanalytic concept of countertransference (Freud, 1910, 1912). It has been defined as the result of the patient's influence on the analyst's unconscious feelings or, in other words, the analyst's transference on the patient. Stemming from unresolved psychological conflicts of the analyst, countertransference was originally considered an obstacle to the treatment of the patient. Later, in the 1950s, this overly restrictive perspective underwent a radical revision (cf., Heimann, 1950). According to the totalistic approach (Kernberg, 1965), countertransference was viewed as the wide range of feelings, thoughts, attitudes and behaviors experienced by the clinician in treating the patient; in these terms, it could be considered a valuable source of information about the patient's intrapsychic and interpersonal dynamics. Therefore, the clinician reactions to a patient impact the diagnostic and therapeutic process, promoting a more accurate understanding of the patient's psychological functioning, especially in the treatment of personality disorders (Beck et al., 2004; Dahl et al., 2014; Gabbard, 2014; Lingardi and McWilliams, 2015; Yeomans et al., 2015; Bateman and Fonagy, 2016).

By definition, personality disorders are dysfunctional schemas of the self and the relationship between self and others. Patients' styles of relating often emerge in the clinical relationship, when the therapist is drawn into interactions that reflect the patient's enduring and maladaptive psychological and interpersonal dynamics (Bateman and Fonagy, 2006; Clarkin et al., 2006; Gabbard, 2014; Lingardi and McWilliams, 2017).

Accordingly, therapists' recognition of their subjective reactions to a patient is important for their deep understanding of the patient's relational patterns and inner experience.

Research has examined the association between therapists' responses and patients' personality syndromes, especially from the perspectives of clinicians and external observers. Some studies, based on the clinician's perspective, have evaluated the quality and intensity of therapists' emotional responses to patients presenting with DSM-IV (American Psychiatric Association [APA], 1994) axis II clusters (e.g., Betan et al., 2005; Rössberg et al., 2007), specific personality disorders (e.g., Bourke and Grenyer, 2013; Colli et al., 2014; Lingardi et al., 2015; Tanzilli et al., 2017; Tanzilli and Gualco, 2020), and personality traits in the psychotherapy context (e.g., Rössberg et al., 2008; Tanzilli et al., 2018). Other empirical investigations, based on the observer's perspective, have assessed therapists' emotional experiences using vignettes or audio/video recordings of personality disordered patients (e.g., McIntyre and Schwartz, 1998; Schwartz et al., 2007). All of these studies have shown that cluster A disorders (i.e., paranoid, schizoid, and schizotypal personality disorders) are associated with disengaged response pattern; cluster B disorders (i.e., antisocial, borderline, histrionic, and narcissistic personality disorders) are correlated with overwhelmed/disorganized feelings, helplessness, hostility, withdrawal, and sexual attraction; and cluster C disorders (i.e., avoidant, dependent, and obsessive–compulsive personality disorders) are associated with nurturant and warm feelings. Moreover, these results have been consistent across therapists' theoretical orientations. Overall, there is an increasing consensus that therapists' awareness of their personal responses to patients may promote greater sensitivity in the diagnostic process, more accurate and clinically meaningful case formulations, and more effective therapeutic interventions (e.g., Gabbard, 2009a; Bateman and Fonagy, 2016; Lingardi and McWilliams, 2017).

The present study adopted the totalistic approach (Kernberg, 1965) to explore which specific therapist reactions and neural responses allow clinicians to discriminate between patients with different personality disorders. To the best of our knowledge, no empirical research has ever examined

therapists' response patterns to patients from a neuroscience perspective. In addition, considering that previous studies have reported the influence of analyst gender on transference and countertransference responses (Berg et al., 2019; Chertoff et al., 2020), to prevent this effect from influencing the results, only male therapists were included in the present study. Specifically, the research aimed at contributing a preliminary empirical investigation of the clinical and neural responses—evaluated in terms of event-related potential (ERP) components—of psychodynamic clinicians while watching selected still images from videos of three patients (performed by actors) with various personality disorders (i.e., narcissistic, histrionic/borderline, depressive).

Event-related potentials (ERPs) are cortical measurements of the total electrical activity of postsynaptic potentials, which are generated in brain structures in response to sensory, cognitive, and motor events or stimuli (Peterson et al., 1995; Ghani et al., 2020). ERPs are collected *via* a non-invasive procedure, and they enable an advanced study of the temporal dynamics of stimulus processing (Luck, 2014). The components of the averaged ERP waveform indicate deflection (i.e., P for positive, N for negative), expected latency from the stimulus onset, and amplitude (i.e., neural resources required for processing) (Luck, 2005). ERPs that peak within the first 100 ms (approximately) after stimulus onset are known as sensory or exogenous, as they depend largely on the physical parameters of the stimulus; in contrast, later ERPs (emerging after 100 ms) are termed cognitive or endogenous, as they reflect the evaluation of information processing (Sur and Sinha, 2009; Sokhadze et al., 2017). Moreover, while early ERPs (i.e., N100, N200, P200) are mainly implicated in attention selection processes, subsequent ERPs (i.e., P300 or later) reflect stimuli evaluation or categorization.

The P1 component reflects early sensory processing within the extrastriate visual cortex (Vogel and Luck, 2000; Olofsson et al., 2008). It is typically larger in response to emotional stimuli (Carretié et al., 2004)—particularly faces (Holmes et al., 2007; Mueller et al., 2009; Mühlberger et al., 2009)—than non-emotional stimuli. The subsequent N100 component is a centroparietal negative deflection, which peaks at approximately 130 ms after stimulus onset (Keil et al., 2001; Foti et al., 2009). Relative to P1, N1 reflects increased sensitivity to the emotional content of a visual stimulus, as it is larger for both pleasant and unpleasant stimuli, relative to neutral pictures (Keil et al., 2001; Carretié et al., 2007; Foti et al., 2009; Weinberg and Hajcak, 2011). P200 peaks at approximately 180 ms after stimulus onset (Carretié et al., 2004), and it is maximal at the anterior and central sites (Luck and Hillyard, 1994). This ERP component reflects the early affective evaluation and discrimination of visual stimuli (Begleiter et al., 1979; Conley et al., 1999) such as facial expressions (Eimer et al., 2003), as well as emotional words (Kissler et al., 2006; Kanske and Kotz, 2007). N2 is a fronto-central negativity that follows P2 and

peaks at approximately 200–350 ms after stimulus onset. It is thought to be generated in the anterior cingulate cortex (ACC; Falkenstein et al., 1999; Pires et al., 2014). N2 is considered an index of cognitive control processes, as it deals with the inhibition of incorrect responses and is larger for conflict resolution tasks (Sokhadze et al., 2017; Ligeza et al., 2018). The subsequent broad centro-parietal positive deflection (i.e., P3) occurs between 300 and 600 ms after stimulus onset (Roche et al., 2004; Polich, 2007), and is thought to be generated by a more distributed network of cortical regions, relative to N2 (Polich, 2007; Foti and Hajcak, 2008; MacNamara and Hajcak, 2009, 2010). P3 amplitude is sensitive to motivationally salient stimuli, as it is modulated by both pleasant and unpleasant cues, regardless of whether the salience is defined in terms of the task demand or stimulus content (Olofsson et al., 2008; Hajcak et al., 2010; Weinberg and Hajcak, 2010); similar results have been reported for emotional words (Naumann et al., 1992) and faces (Allison et al., 1999; Grecucci et al., 2018; Sulpizio et al., 2019). Finally, the late positive potential (LPP) is a sustained positive deflection in the ERP waveform with centroparietal distribution, which emerges at approximately 300 ms following stimulus onset and persists for the duration of the stimulus (Schupp et al., 2000, 2004a; Foti et al., 2009) and beyond (for as long as 1,000 ms) (Hajcak and Olvet, 2008; Hajcak et al., 2010). LPP amplitude is larger for emotionally evocative stimuli (i.e., appetitive and aversive stimuli) than for neutral stimuli (Dunning and Hajcak, 2009; Hajcak et al., 2009), and it covaries with subjective arousal ratings of emotional stimuli (Cuthbert et al., 2000; Weinberg and Hajcak, 2010). Simultaneous functional magnetic resonance imaging–electroencephalogram (fMRI-EEG) studies have shown that LPP amplitude is also associated with activity in posterior cerebral regions (especially, lateral occipital, inferotemporal, and parietal visual areas), which are implicated in attention to and the perceptual processing of the motivational relevance of visual scenes (Sabatinelli et al., 2007). In the present research, three subcomponents of the LPP—hereafter termed late components (LC)—were identified according to their persistence: LC1 (i.e., 400–500 ms), LC2 (i.e., 500–700 ms), and LC3 (i.e., 700 ms).

In this study, supervised machine learning (ML)—known as a decision tree (Pekel and Özmen, 2020)—was applied (for the first time) to the domain of therapist responses. The advantage of a supervised ML approach (i.e., a branch of artificial intelligence) over standard frequentist approaches is that the algorithm extracts a mathematical function that maps one variable to another (e.g., ERP components to patients' personality disorders), in order to predict new cases (Pekel and Özmen, 2020; Dadomo et al., 2022; Grecucci et al., 2022). Indeed, ML model performance refers to the prediction accuracy for new observations, rather than the degree to which certain factors explain the data.

Based on the above considerations, this pilot research aimed at evaluating whether: (1) therapists' responses predicted which

patient they observed and whether particular clinician reactions contributed to distinguishing between patients with different personality disorders; and (2) therapists' ERPs predicted which patient they observed and, in particular, whether distinct ERP components allowed this forecast.

Materials and methods

Participants

The present study was conducted at the Department of Dynamic and Clinical Psychology, and Health Studies, Faculty of Medicine and Psychology, Sapienza University of Rome, in compliance with the Declaration of Helsinki (Helsinki, Finland, June 1964). A sample of 14 licensed therapists ($M = 36.07$; $SD = 2.97$; range 31–40) was recruited according to the following inclusion criteria: (a) self-identification as a cisgender man; (b) aged 30–40 years; and (c) reporting normal or corrected-to-normal vision. The exclusion criteria were: (a) the presence of neurological injury and psychiatric disease; (b) habitual drug or alcohol use; and (c) visual impairment. Participants self-declared their absence of psychiatric and neurological disease and use of drugs and alcohol. All clinicians provided written informed consent to participate and received no remuneration. The study was approved by the Department Research Ethics Committee (protocol number: 0000112/2020; date: 20/12/2020).

Measures

Clinical questionnaire. It is an *ad hoc* clinician-report questionnaire that was used to obtain information about the therapists. Clinicians provided basic demographic and professional data, including their years of clinical experience.

Shedler-westen assessment procedure-200 (SWAP-200; Westen and Shedler, 1999a,b; Shedler et al., 2014). The SWAP-200 is a psychometric system designed to provide a comprehensive assessment of personality and personality pathology. It consists of 200 items that clinicians sort into eight categories, ranging from 0 (not descriptive of the person) to 7 (most descriptive of the person), in order to comply with a fixed distribution. The SWAP-200 assessment furnishes: (a) a personality diagnosis, based on the matching of the patient assessment with 10 personality prototypes from the DSM-IV axis II (i.e., PD scales); (b) a personality diagnosis, based on the matching of the patient's SWAP description with 11 empirically derived Q-factors/styles of personality; and (c) a dimensional diagnostic approach, based on a multifaceted model of personality pathology, including 12 clinically relevant dimensions [e.g., hostility, narcissism, emotional dysregulation, dysphoria, schizoid orientation;

see Shedler and Westen (2004)]. The tool also includes a dimensional measure of psychological strengths and adaptive functioning. A personality disorder is assigned when scores on one or more PD scales and/or Q-factor or personality traits (in standardized T-scores) are ≥ 60 and the high-functioning scale score is < 60 . The present study used only the SWAP-200 PD scales and personality dimensions, along with the healthy personality global functioning index (i.e., high-functioning scale). The SWAP-200 was designed for use by clinically experienced informants, and its reliability and validity have been extensively tested on different patient populations in several studies, including multi-observer studies comparing diagnoses by treating therapists with diagnoses by independent assessors, based on research interviews (e.g., Westen and Muderrisoglu, 2006; Blagov et al., 2012).

Therapist response questionnaire (TRQ; Zittel Conklin and Westen, 2003; Betan et al., 2005). The TRQ is a clinician-report instrument that assesses therapists' emotional responses to patients. It consists of 79 items that measure a wide spectrum of thoughts, feelings, and behaviors expressed by therapists toward patients. Clinicians evaluate each item on a 5-point Likert scale ranging from 1 (not true) to 5 (very true). The present study used an empirically supported version of the TRQ (Tanzilli et al., 2016) to evaluate nine therapist response patterns: (a) helpless/inadequate, describing feelings of inadequacy, incompetence, and inefficacy; (b) overwhelmed/disorganized, describing intense feelings of overwhelm in response to the patient's emotions and needs, as well as confusion, anxiety, or repulsion; (c) positive/satisfying, describing an experience of close connection, trust, and collaboration with the patient; (d) hostile/angry, describing feelings of anger, hostility, and irritation toward the patient; (e) criticized/devalued, describing a sense of being criticized, dismissed, or devalued by the patient; (f) parental/protective, describing a wish to protect and nurture the patient in a parental manner; (g) special/overinvolved, describing that the patient is very special, to the extent that the clinician may show some difficulty maintaining the boundaries of the therapeutic setting; (h) sexualized, describing the presence of sexual attraction toward the patient; and (i) disengaged, describing feelings of annoyance, boredom, withdrawal, or distraction during sessions. Scores for each scale are obtained by calculating the average score of the items comprising each factor. The nine TRQ factors showed excellent internal consistency (Streiner, 2003), obtaining the following Cronbach's alpha values: criticized/devalued ($\alpha = 0.86$), helpless/inadequate ($\alpha = 0.91$), positive/satisfying ($\alpha = 0.84$), parental/protective ($\alpha = 0.77$), overwhelmed/disorganized ($\alpha = 0.88$), special/overinvolved ($\alpha = 0.78$), sexualized ($\alpha = 0.87$), disengaged ($\alpha = 0.80$), and hostile/angry ($\alpha = 0.87$). The TRQ (and its version for adolescents, TRQ-A; Tanzilli et al., 2020) showed high reliability and validity in different clinical populations (cf., Stefana et al., 2020).

Materials

Patient videos

Three videos were derived from the psychotherapy sessions depicted in the popular American television series *In Treatment*. Specifically, the videos showed clinical interactions between the therapist, Paul Weston, and three of his patients: Alex Prince, Laura Hill, and April (without surname). These three characters were chosen because they presented different personality pathologies (i.e., narcissistic, histrionic/borderline, depressive), and all of the clinical interactions between Paul (the therapist, played by an actor) and the three patients (performed by actors) clearly showed crucial aspects of psychological and interpersonal functioning related to these specific personality syndromes.

The selected contents of all three videos are reported below.

Alex

Alex is a military aircraft pilot. The video shows his initial exchange with Paul during their first psychotherapy session: Alex enters Paul's office without introducing himself. He repeatedly asks whether Paul recognizes him, without explaining why. Alex tells Paul that he has collected information about Paul's professional skills, and ultimately chose Paul because the data confirmed that "he was the best." Alex reveals that he is accustomed to engaging with only the best. Thus, right from the first lines spoken, Alex shows an exaggerated sense of self-importance. He also appears dismissive and devaluing of Paul, interrupting him often and accusing him of not being a "good listener." Alex structures and tries to dominate the interaction with Paul, leaving little room for the therapist, who feels cornered. Alex reveals that he is a war hero, and for that reason, he expected Paul to recognize him. In fact, he is the pilot who completed a delicate mission in Iraq that, by mistake, resulted in the death of many children. In the face of Paul's disbelief, Alex emphasizes that he does not feel guilty. Rather, he proudly affirms that he did his duty with surgical precision, confirming the standards of excellence he had always guaranteed as one of the most qualified 'top guns' in a Navy special department.

Laura

Laura is an anesthesiologist who has been in a long-term relationship with another one of Paul's patients, which seems to be progressing toward marriage. However, she is overwhelmed by her attraction to Paul. The clip begins in the middle of a conversation between Laura and Paul: she has just confessed her love to Paul and is openly disappointed by his reaction, having hoped that he would have felt the same. Laura details romantic and erotic fantasies about Paul's positive response, trying to provoke his reaction. She also expresses frustration and anger toward Paul, pointing out how humiliating it is to see him do nothing after her declaration of love, and showing

difficulty accepting rejection. At one point, Laura talks about a sexual encounter she had with a stranger in a club. She says that the reason she did not have sex with that man was not because she wished to remain faithful to her boyfriend, but because she was deeply in love with Paul. Continuing to provoke him, she describes in detail her fantasy of meeting Paul in a bar and being seduced by him. She then states that this is only her imagination. She knows nothing real can ever happen between them; however, she cannot resign herself to this reality. At the end of the clip, Laura appears desperate and lamenting, wondering what will become of her life and how she will manage this situation.

April

April is a young architecture student who has recently been diagnosed with lymphoma. In the clip, she never talks about her illness, but instead speaks about how she feels and how she has always felt in her life. April appears tired. Something about her eyes seems off and she feels depleted of strength. April tells Paul that she did not sleep a wink the previous night. Although she tries to make some jokes, her worry and sadness are evident. She describes her sleepless night as a nightmare and reveals that she sometimes considers hurting herself in order to end her excruciating pain. However, she quickly reassures Paul by saying that she is not serious, and that she would never do something so reckless. She is just trying to share her feelings with him. April talks about her fear of going crazy, recalling that she had this concern even when she was younger. However, she refuses to explore the topic further. She relates to Paul the thoughts that were troubling her the night before, and expresses—with a wistful attitude—her desire to return to a faraway place where she once felt safe. Paul asks her to describe this place in more detail, but April does not respond, and instead slumps down on the couch. She tells Paul to wake her up a little later, and then falls asleep.

Patient personality profiles

The personality syndromes of the abovementioned characters were assessed using the SWAP-200 (Westen and Shedler, 1999a; Shedler and Westen, 2004; Shedler et al., 2014), by two independent judges who blindly viewed the first three treatment sessions of each patient. The inter-rater reliability (IRR) was good (Spearman's $Rho = 0.75$). Figures 1, 2 show the personality profiles of the three patients according to the SWAP-200 PD and personality dimension scales, respectively.

Considering the SWAP-200 PD scales, Alex presented with narcissistic personality disorder (T-score = 71), alongside clinically relevant antisocial and paranoid traits (T-scores = 59.8 and 56.7, respectively). His high-functioning score demonstrated an average level of overall personality functioning (T-score = 53.3). Laura presented with histrionic (T-score = 76.6) and borderline (T-score = 68.4) personality disorders, as well as strong dependent personality traits (T-score = 56.6). Her high-functioning score was slightly below

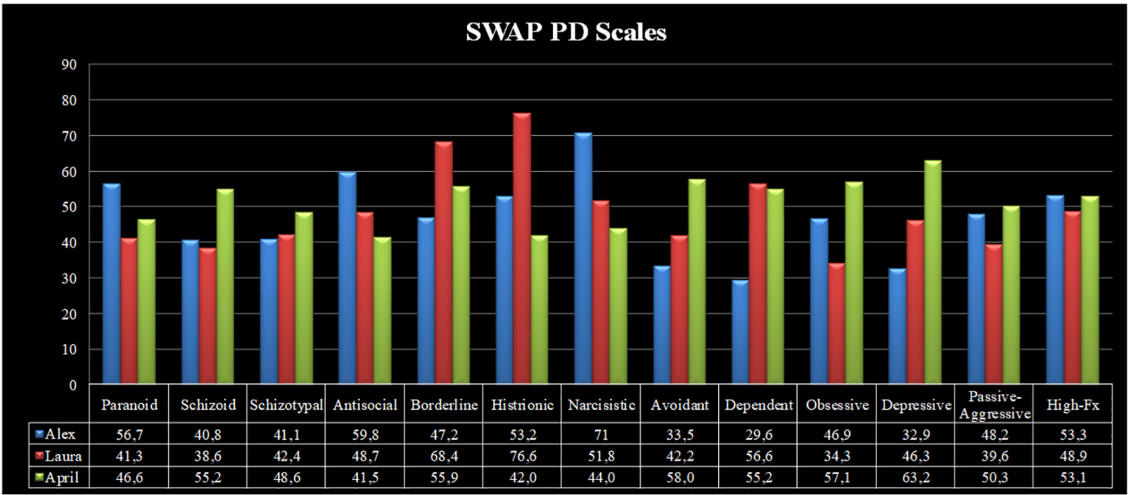


FIGURE 1
SWAP-200 personality profiles of Alex, Laura, and April.

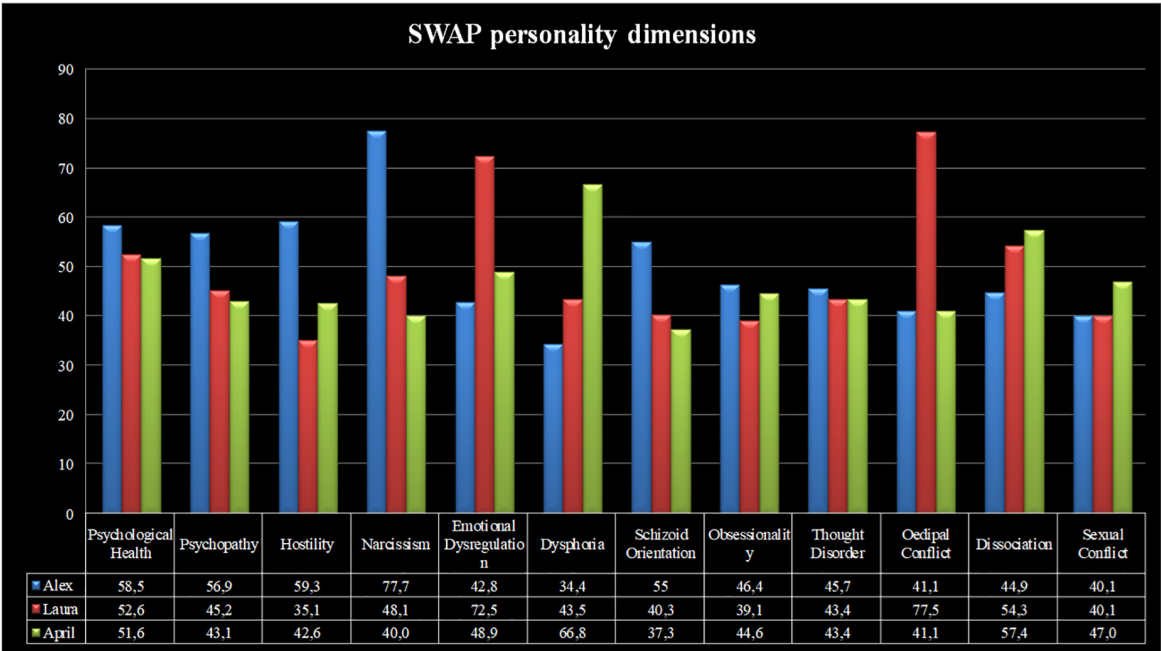


FIGURE 2
SWAP-200 personality dimensions of Alex, Laura, and April.

average (T-score = 48.9). Finally, April showed a depressive personality disorder (T-score = 63.2), in addition to clinically relevant avoidant (T-score = 58) and obsessive (T-score = 57.1) personality traits. Her personality functioning was average (T-score = 53.1).

In terms of the SWAP-200 personality dimensions, Alex presented with high levels of narcissism (T-score = 77.7), along with clinically meaningful traits of hostility (T-score = 59.3) and

psychopathy (T-score = 56.9). Laura showed a great degree of Oedipal conflict (T-score = 77.5) and emotional dysregulation (T-score = 72.5), whereas April presented with high levels of dysphoria (T-score = 66.8).

Visual stimuli

The visual stimuli were comprised of 72 images of faces in color, with neutral emotional valence and a similar oval

shape. Thirty-six still images depicting the patients' faces (i.e., target stimuli; 12 for Alex, 12 for April, and 12 for Laura, respectively) were selected from the previously shown videos. Additionally, 36 images depicting unfamiliar faces (i.e., filler stimuli) were selected from the racially diverse affective expression (RADIATE) stimulus set (Conley et al., 2018), on the basis of their similarity to the three patients: 12 unfamiliar faces of African American men were chosen as Alex's distractors, 12 unfamiliar faces of White women were chosen as April's distractors, and 12 unfamiliar faces of White women were chosen as Laura's distractors. During the visual stimuli task, the 72 images were randomly repeated for three times to get a total of 216 trials [36 trials *per* stimulus (Alex, April, Laura, distractor of Alex, distractor of April, and distractor of Laura)]. Each trial started with a fixation cross displayed for 1,000 ms, followed by the visual stimulus (Alex, April, Laura, distractor of Alex, distractor of April, and distractor of Laura) presented for 2,000 ms. The trial ended with inter-stimulus interval (ISI) of 500 ms (Figure 3). The visual stimuli task lasted about 13 min. Given the specific experimental question (i.e., "Can we classify one patient versus the other?"), this paper does not report the analyses of filler stimuli. Rather, the present study focused on the faces of the real patients, to determine whether ERP components and therapist reactions could discriminate (i.e., classify, in ML terms) between each pair of patients (i.e., Alex vs. April, Alex vs. Laura, and April vs. Laura).

Experimental procedure

Therapists were asked to complete a battery of instruments (including the clinical questionnaire and the TRQ; see "Measures" section) after watching the three video clips described above. Each participant filled out the questionnaires in consideration of how they might feel when interacting with the patients observed in the videos. To prevent any order effect, videos were shown to participants in a random order. Subsequently, clinicians were asked to complete a laboratory experiment in which they viewed a series of still images from the videos. All participants performed the visual task in a dimly lit room, seated at a distance of 80 cm from the PC monitor displaying the stimuli (27 cm, 75 Hz, 1,024 × 768). The visual stimuli were presented using E-Prime software (v.2.0.8.90; Psychology Software Tools), and the high-density electroencephalography (hdEEG) signal of each participant was recorded during the task. To limit potential biases associated with preconceived stereotypes about personality syndromes, participants were unaware of patients' personality diagnoses. Following the experimental procedure, clinicians were asked to assign a personality diagnosis to all three patients based on the observed video material. The reliability of their responses was about 90%.

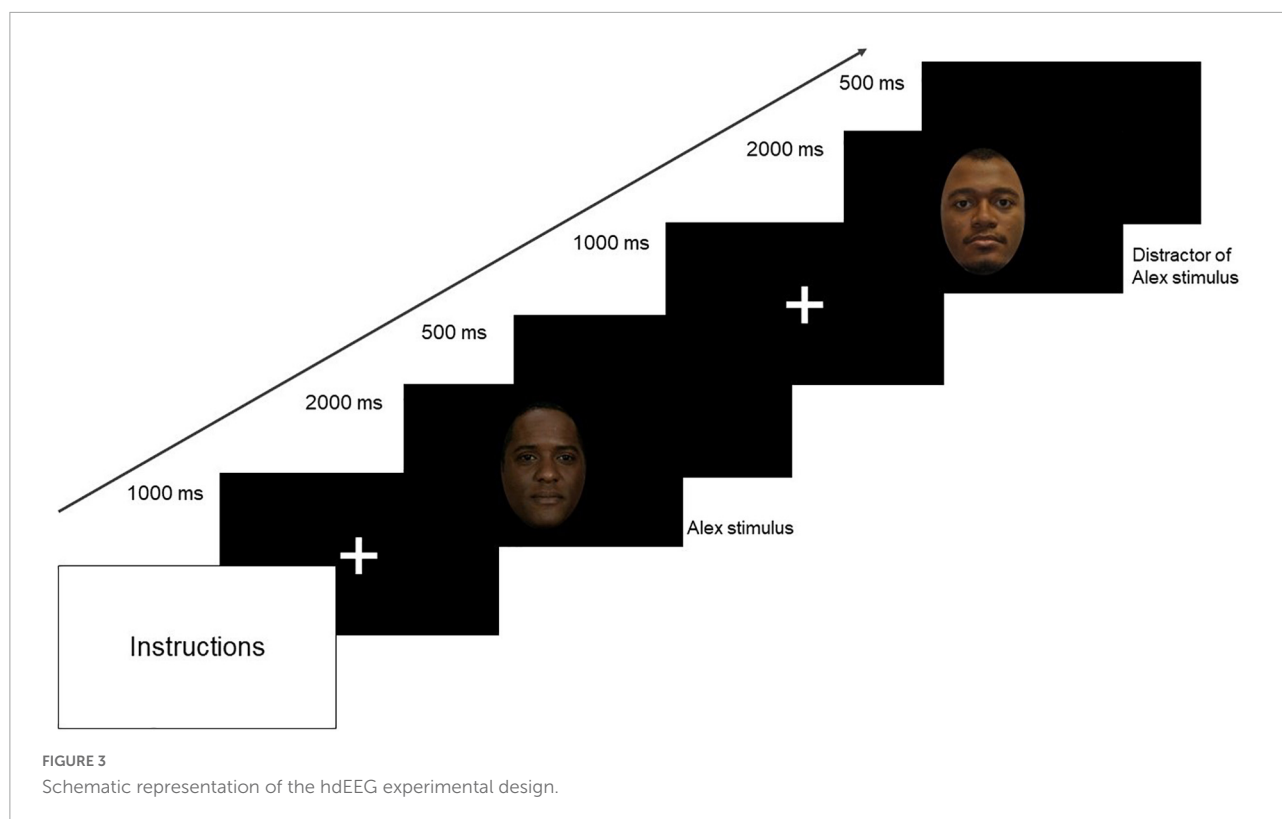
High-density electroencephalography recordings and data processing

HdEEG signals were recorded continuously at 250 Hz with reference to the vertex (Cz), with impedances kept below 50 k Ω , using a 256-channel Hydrocel Geodesic Sensor Net and Net Station software (v.4.4.2; Electrical Geodesic, Inc., Eugene, OR, USA). Subsequently, hdEEG data were digitally filtered at 30 Hz low-pass in offline mode. Data for each participant were segmented in epochs of 1,100 ms, ranging from 100 ms before to 1,000 ms after stimulus onset. Artifact detection was set to 200 μ V for bad channels (i.e., noisy electrodes), 140 μ V for eye blinks, and 100 μ V for electrodes detecting eye movements (Picton et al., 2000; Bourisly and Shuaib, 2018; Lai et al., 2020; Altavilla et al., 2021). Segments containing eye blinks, eye movement, or more than 15 bad channels were excluded. A baseline correction of −100 ms before stimulus onset was applied. The amplitude and the latency of ERP components in response to each type of stimulus (Alex, April, and Laura) were extracted automatically through Net Station software, averaging the clean segmented trials for each participant. In the study, the following intervals were set: 80–160 ms for P100; 160–220 ms for N170; 270–400 ms for P300; 400–500 for LC1; 500–700 for LC2; and 700–1,000 for LC3. Following hdEEG signal cleaning from artifacts, as reported in previous studies (Picton et al., 2000; Tanner et al., 2015; Lai et al., 2020; Altavilla et al., 2021), the following electrode locations were chosen for each montage: occipital (O1 and O2), occipito-temporal (left: 85, 96, 107, 108, 109; right: 151, 160, 161, 171, 172), parietal (left: 77, 78, 79, 86, 87, 88; right: 142, 143, 153, 154, 162, 163), and frontal (left: 23, 24, 30, 35, 42; right: 6, 7, 206, 215, 224).

Data were analyzed for peak amplitudes and latencies at P100 on the occipital and occipito-temporal montages; and at N170 on the occipital, occipito-temporal, and parietal montages. Moreover, data were analyzed for mean amplitudes and latencies at P300 on the occipito-temporal, parietal, and frontal montages. Lastly, at LC1, LC2, and LC3, data were analyzed for mean amplitudes of the occipito-temporal, parietal, and frontal montages.

Source analysis (sLORETA)

To locate the neural generators of the ERP components, hdEEG signals were processed using the standardized low-resolution electromagnetic tomography (sLORETA) (Pascual-Marqui, 2002) inverse model of the GeoSource software (v.2.0; EGI, Eugene, OR, USA), with Tikhonov 1×10^{-4} regularization. sLORETA assumes the current density standardization, which considers not only the variance of noise in the hdEEG measurements but also the biological variance in the actual signal (Pascual-Marqui, 2002). Biological variance is thought to



be independent and uniformly distributed in the brain, resulting in a linear image localization with exact and zero localization error (Jatoi et al., 2014). Source locations were derived from the probabilistic map of the Montreal Neurological Institute 305 subjects (i.e., MNI305 average). On this basis, gray matter volume was parcellated into 7-mm voxels. Each voxel served as a source location with three orthogonal orientation vectors, resulting in a total of 2,447 source triplets, whose anatomical labels were estimated using a Talairach Daemon (Lancaster et al., 2000; Cecchini et al., 2013; Massaro et al., 2018). Magnetic resonance imaging normalization and data extraction were performed for each participant, and the mean intensities of specific Brodmann areas (BAs) were extracted for each ERP component.

In accordance with the hypotheses and the literature regarding the neurobiological correlates of the visual processing of faces (Eimer and Holmes, 2007; Vuilleumier and Pourtois, 2007; Hung et al., 2010; Lai et al., 2020), the following regions of interest (ROIs) were chosen for the sLORETA analyses: occipital, limbic, anterior cingulate cortex, posterior cingulate cortex, parietal, temporal, and prefrontal cortex. For each ROI, the following BAs in both hemispheres were selected: BA17, BA18, and BA19 for the occipital ROI; amygdala, insula, amygdala-hippocampus junction, and hippocampus for the limbic ROI; BA24, BA32, and BA33 for the anterior cingulate cortex ROI; BA23, BA30, and BA31 for the posterior cingulate cortex ROI; BA01, BA02, BA03, BA05, and BA07 for the parietal

ROI; BA20, BA 21, BA 22, BA 37, BA 38, BA 41, BA 42, and BA 43 for the temporal ROI; and BA09, BA10, BA11, BA46, and BA47 for the prefrontal cortex ROI. The mean intensity of each BA in response to the target stimuli (i.e., Alex, April, Laura) was extracted for each ERP component (i.e., P100, N170, P300, LC1, LC2, LC3).

Statistical analysis

Statistical analyses were performed in JASP (v.016; JASP Team 2021), using a decision tree classification model. The decision tree is a supervised ML algorithm that obtains predictive estimates for variables that take discrete values (Loh, 2011). Tree models in which the target variable has a discrete value are called classification trees. Such trees are structured so that every leaf represents a class label (i.e., Alex, Laura, April), and branches represent relevant features that predict class labels (i.e., therapist reaction, ERP components). The algorithm finds the optimal decision tree by computing the error between the predicted value and the actual value at each split point. Split point errors are then compared across variables and the lowest prediction error is used to generate the tree (Pekel and Özmen, 2020). Such split points indicate the most relevant predictive features.

In the present study, the data split was set to use 20% of the data: of the 28 observations (i.e., 14 therapist responses for

two patients at a time) 23 were used for model training and five were used for model testing. The hold-out method was used to test the predictive value of each model, with the following split: 80% of observations were used to train the model, and the remaining 20% were used to test the model (i.e., 23 observations were used for training and five were used for testing). The algorithmic settings were as follows: the minimum number of observations for a split was set to 20, the minimum number of observations in the terminal was set to 7, the max iteration depth was set to 30; and the complexity parameter was set to 0.01. Predictors were scaled. Models were iterated five times and the best model was selected on the basis of the highest accuracy (i.e., bake-off method).

Results

Characteristics of therapists' responses to patients

Figure 4 depicts the therapist responses to patients with narcissistic, histrionic/borderline, and depressive personality disorders (i.e., Alex, Laura, and April, respectively). The mean scores of the TRQ patterns indicated that the strongest therapist response patterns elicited by Alex were criticized/devalued ($M = 3.30$, $SD = 0.62$) and hostile/angry ($M = 3.17$, $SD = 0.87$), whereas the most intense therapist reactions patterns evoked by Laura and April were sexualized ($M = 3.02$, $SD = 0.97$) and parental/protective ($M = 3.06$, $SD = 0.97$), respectively.

Prediction of patient personality disorder from therapists' responses and ERP components

The results of the decision tree are reported below for each pair of patients.

The first comparison focused on the therapist reactions to the narcissistic (i.e., Alex) versus the depressive (i.e., April) patient. The decision tree performance when classifying Alex versus April achieved a classification accuracy of 100%. Of note, the precision (i.e., positive cases predicted) was 100% and the recall (i.e., true positive rate) was 100%. The area under the curve was 1.0.

The relative importance of all therapist response patterns included in the ML model were, in order: criticized/devalued (26.190), hostile/angry (19.048), helpless/inadequate (16.667), parental/protective (14.286), special/overinvolved (11.905), and disengaged (11.905) (**Figure 5A**). Of note, the criticized/devalued pattern was responsible for the main split between Alex and April, according to the following rule: if criticized/devalued ≥ 0.065 then Alex, if criticized/devalued < 0.065 then April.

In the comparison focused on ERP components, the decision tree performance when classifying Alex versus April achieved a classification accuracy of 80%. Thus, the extracted function was able to accurately guess which patient the therapist was observing at a level far above chance. The precision (i.e., positive cases predicted) was 85% and the recall (i.e., true positive rate) was 80%. The area under the curve was 0.75%.

The relative importance of ERP components was, in order: LC1 RHPC (19.149), LC2 RHPC (19.149), LC3 LHPC (17.021), LC2 LHPC (17.021), P1 RAMG (14.894), and LatP3 frL (12.766) (**Figure 5B**). Of note, LC1 RHPC was responsible for the main split between Alex and April, according to the following rule: if LC1 RHPC ≥ 0.099 then Alex, if LC1 RHPC < 0.099 then April.

The second comparison focused on the therapist responses to the narcissistic (i.e., Alex) versus the histrionic/borderline (i.e., Laura) patient. The decision tree performance when classifying Alex versus Laura achieved a classification accuracy of 80%. The precision (i.e., positive cases predicted) was 87% and the recall (i.e., true positive rate) was 80%. The area under the curve was 0.83.

The relative importance of therapist responses was, in order: criticized/devalued (29.412), hostile/angry (20.588), helpless/inadequate (11.765), positive/satisfying (11.765), overwhelmed/disorganized (11.905), and parental/protective (8.824) (**Figure 6A**). Of note, the criticized/devalued pattern was responsible for the main split between Alex and Laura, according to the following rule: if criticized ≥ 0.079 then Alex, if criticized < 0.079 then Laura.

In the comparison focused on ERP components, the decision tree performance when classifying Alex versus Laura achieved a classification accuracy of 80%. Thus, the extracted function was able to accurately guess which patient the therapist was observing at a level far above chance. The precision (i.e., positive cases predicted) was 85% and the recall (i.e., true positive rate) was 80%. The area under the curve was 0.75.

The relative importance of ERP components was, in order of importance: LC2 RHPC (21.277), LC3 RHPC (19.149), LC1 RHPC (19.149), LC2 LHPC (17.021), LatP3 frL (12.766), and N1 RAMG (10.638) (**Figure 6B**). Of note, LC2 RHPC was responsible for the main split between Alex and Laura, according to the following rule: if LC2 RHPC ≥ 0.331 then Alex, if LC3 < 0.331 then Laura.

The third comparison focused on the therapist reactions to the depressive (i.e., April) versus the histrionic/borderline (i.e., Laura) patient. The decision tree performance when classifying Alex versus April achieved a classification accuracy of 100%. The precision (i.e., positive cases predicted) was 100% and the recall (i.e., true positive rate) was 100%. The area under the curve was 1.0.

The relative importance of therapist reactions was, in order: sexualized (43.478), overwhelmed/disorganized (21.739), parental/protective (13.043), helpless/inadequate

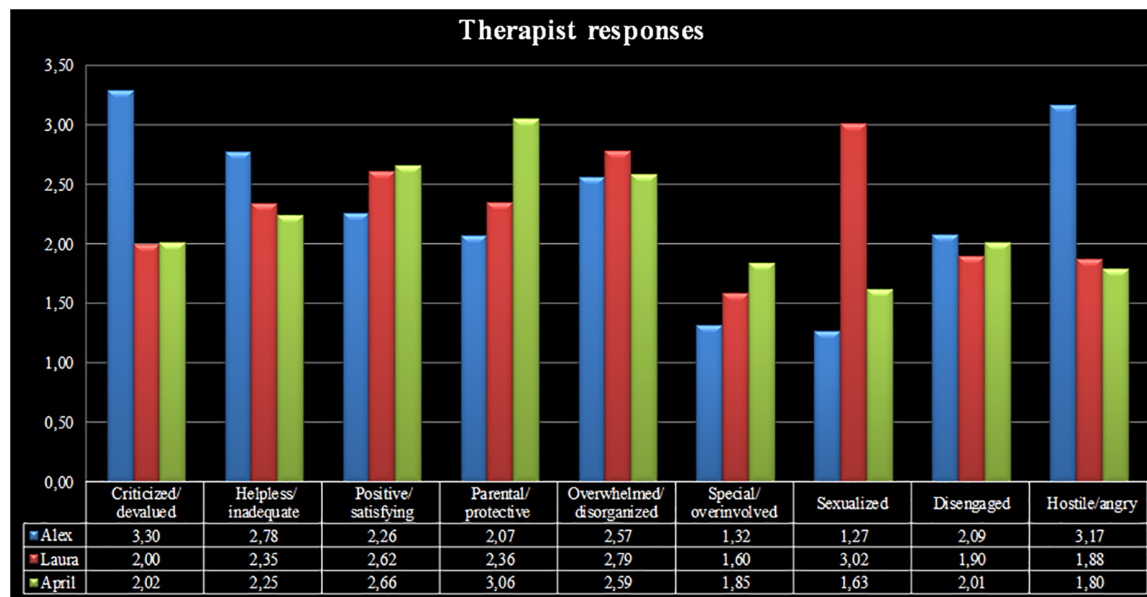


FIGURE 4
Therapist responses to Alex, Laura, and April.

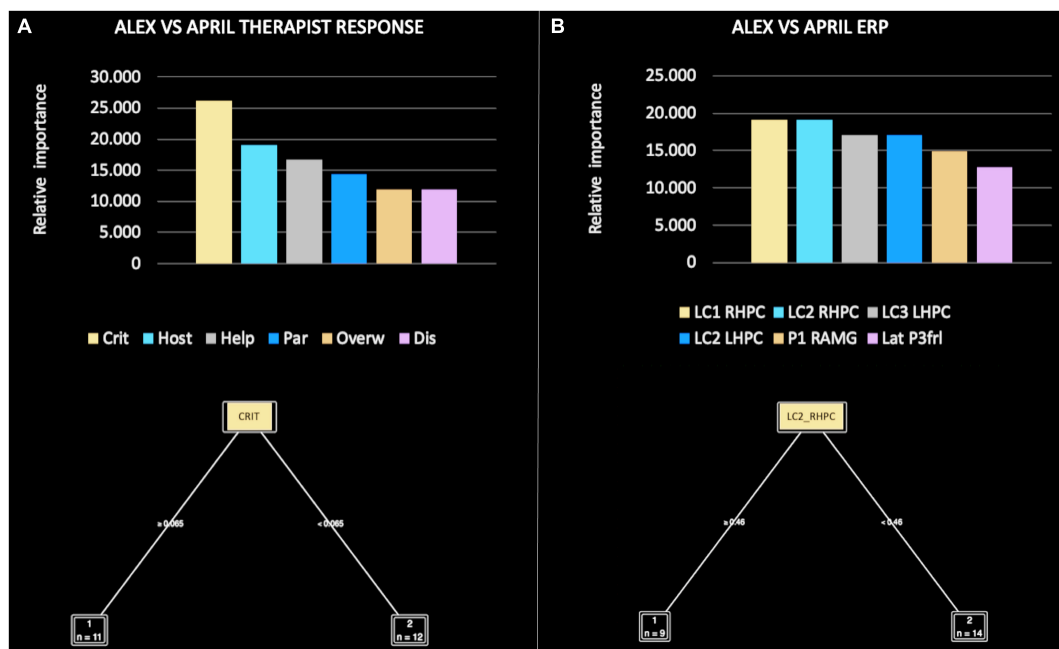


FIGURE 5
Relative importance of all therapist response patterns (A) and event-related potential (ERP) components (B), included in the machine learning (ML) model for Alex versus April.

(8.696), hostile/angry (8.696), and criticized/devalued (4.348) (Figure 7A). Of note, the sexualized pattern was responsible for the main split between April and Laura, according to the following rule: if sexualized ≤ 0.049 then April, if LC3 > 0.049 then Laura.

In the comparison focused on ERP components, the decision tree performance when classifying April versus Laura achieved a classification accuracy of 100%. The precision (i.e., positive cases predicted) was 100% and the recall (i.e., true positive rate) was 100%. The area under the curve was 1.0.

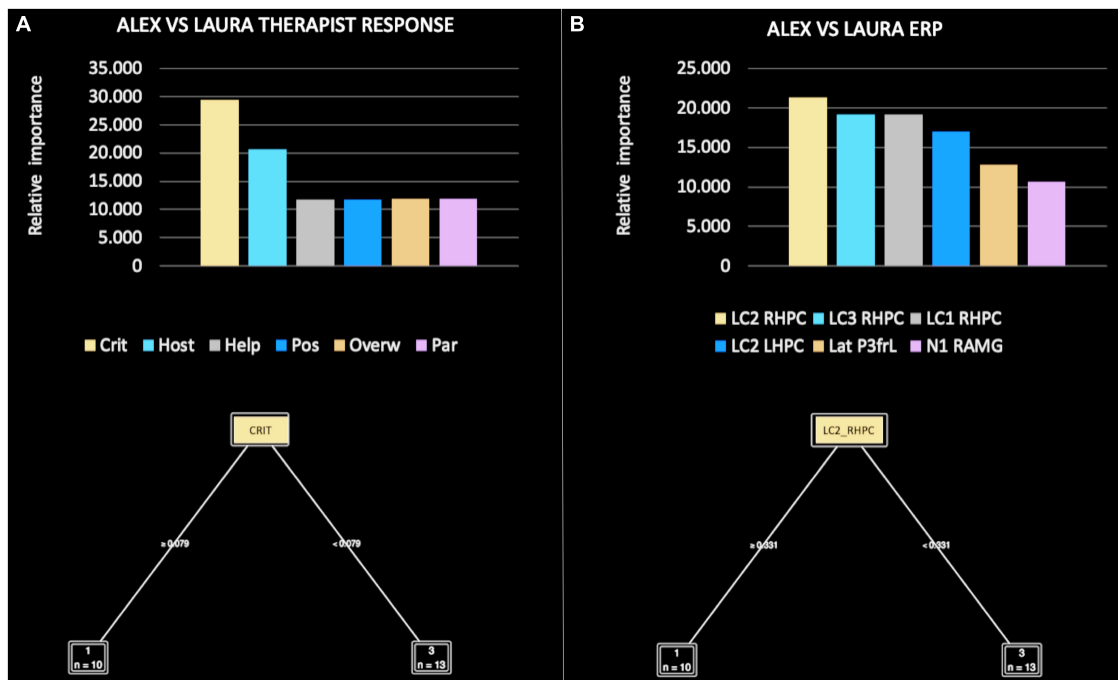


FIGURE 6

Relative importance of all therapist response patterns (A) and event-related potential (ERP) components (B), included in the machine learning (ML) model for Alex versus Laura.

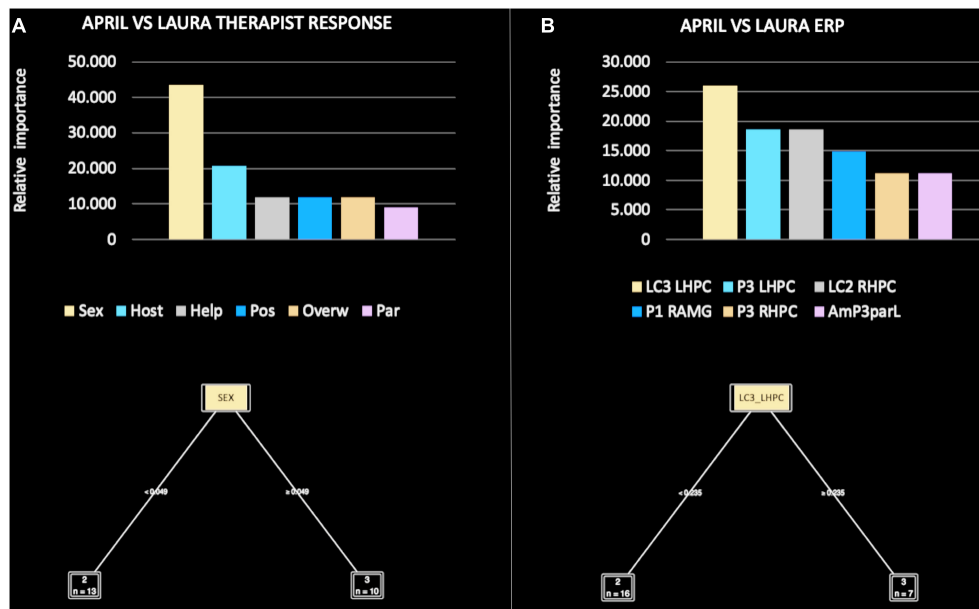


FIGURE 7

Relative importance of all therapist response patterns (A) and event-related potential (ERP) components (B), included in the machine learning (ML) model for April versus Laura.

The relative importance of ERP components was, in order: LC3 LHPC (25.926), P3 LHPC (18.519), LC2 RHPC (18.519), P1 RAMG (14.815), P3 RHPC (11.111), and AmP3 parL (11.111)

(Figure 7B). Of note, LC3 LHPC was responsible for the main split between April and Laura, according to the following rule: if $LC3 \leq 0.235$ then April, if $LC3 > 0.235$ then Laura.

Discussion

The main aims of the present pilot study were to investigate, through an experimental research design: (1) whether therapists' specific emotional, cognitive, motivational, and behavioral responses to visual stimuli depicting three patients with different personality disorders distinguished between patients in a statistically systematic and clinically meaningful way; and (2) whether therapists' distinct neural responses (i.e., ERP components) while viewing visual stimuli of patients' faces during the EEG task differentiated between the three patients, with high accuracy. An ML method was adopted to evaluate the ability of a model (consisting of clinical or neural variables) to accurately predict the facial stimulus of a personality-disordered patient from that of another patient (using pairwise comparisons) during the EEG experimental procedure.

Regarding the first goal, the general question posed was: "Is it possible to predict which patient the therapist observed from their subjective reactions? If so, which pattern of therapist response better predicts one patient than another?" Overall, the findings confirmed that therapists' reactions to visual stimuli depicting patients' faces during the EEG task precisely discriminated between patients presenting with narcissistic (i.e., Alex), histrionic/borderline (i.e., Laura), and depressive (i.e., April) personality disorders. In other words, specific therapist responses were associated with the facial stimuli of patients with particular personality pathologies, in a coherent and predictable manner.

In more detail, the ML models comparing Alex and April (**Figure 5A**), Alex and Laura (**Figure 6A**), and April and Laura (**Figure 7A**) distinguished the visual stimuli of different patients on the basis of clinician reactions with high accuracy (ranging from 80 to 100%). Of note, the criticized/devalued therapist reaction emerged as the most important predictor, with the strongest discriminatory power in identifying the narcissistic patient (i.e., Alex) from the depressive and histrionic/borderline patients (i.e., April and Laura, respectively). Conversely, the sexualized reaction was the most relevant and clinically useful relational dimension in comparing April and Laura, with a cut-off value that maximized both accuracy and the specificity of prediction.

These results are supported by a wide corpus of clinical observations (e.g., Gabbard, 2014; Yeomans et al., 2015; Lingardi and McWilliams, 2017) and empirical evidence in the field (Betan et al., 2005; Dahl et al., 2012, 2014; Bourke and Grenyer, 2013; Colli et al., 2014; Nissen-Lie et al., 2022). In this vein, previous studies investigating the relationship between patients' narcissistic personality disorder and therapists' responses in the psychotherapy context (e.g., Betan et al., 2005; Ronningstad, 2012, 2016; Tanzilli et al., 2017) have highlighted that narcissistic patients typically evoke negative emotional reactions in clinicians, potentially disrupting their ability to benefit from the clinical treatment (McWilliams, 2004).

Clinicians tend to feel devaluated, unappreciated, demeaned, or belittled by narcissistic patients during psychotherapy (e.g., Kernberg, 1975, 2014; Gabbard, 2009b). These reactions may reflect patients' typical affective-interpersonal difficulties, which frequently involve behavior that is domineering, controlling, competitive, hostile, and cold, as well as a defensive tendency to criticize and devalue others due to feelings of inferiority and attempts to stabilize fluctuating self-esteem (e.g., Perry and Perry, 2004; Clemence et al., 2009; Ogrodniczuk and Kealy, 2013).

Other empirical investigations have found that borderline patients tend to elicit therapist response patterns characterized by feelings of overwhelm, helplessness, and overinvolvement (e.g., Lingardi et al., 2015). These reactions may reflect patients' severe emotional dysregulation and contradictory self and other representations, which are related to the overuse of primitive defenses such as spilling and projective identification (e.g., Clarkin et al., 2006). Moreover, histrionic patients tend to evoke a sexualized response in therapists (Gabbard, 2014), perhaps in association with their tendency to display seductive attitudes in their relationships with others (including the clinician). According to the clinical literature (e.g., McWilliams, 2004), this erotization may have a defensive function, serving toward off feelings of weakness or a fear of intimacy in therapy, or to maintain a sense of control and power over the therapist.

Finally, research has shown that depressive patients mainly elicit positive and warm feelings in the therapist (e.g., Blatt, 2004; Blatt and Shahar, 2004; Hennissen et al., 2019; cf. also McWilliams and Shedler, 2017). In general, such patients are cooperative in the therapeutic relationship; accordingly, therapists' nurturant and protective feelings toward depressive patients may relate to their collaborative attitudes. Moreover, depressive patients are characterized by chronic vulnerability to painful affect (especially depression, guilt, shame, and perceived inadequacy), which may provoke in clinicians a strong impulse to care, even to the point of overinvolvement (cf., McWilliams and Shedler, 2017).

Regarding the second goal, the research question was: "Is it possible to predict which patient therapists observed from their brain states (in terms of ERPs)? If so, what are the main ERP components that contribute to this predictive model?" Overall, the findings showed that LPP amplitude in the hippocampus was able to reliably discriminate between patients. The LLP is conceived of as a neural index of the controlled attentional processes (Schupp et al., 2004b; Hajcak et al., 2009; Frenkel and Bar-Haim, 2011) involved in the affective appraisal of stimuli (Schupp et al., 2006; Wessing et al., 2013), and particularly emotional faces.

It has been largely documented that, during facial processing, information from the structural features of the face is rapidly integrated with (and affected by) contextual variables, derived from the observed face (e.g., eye gaze), its surrounding body elements (e.g., body posture), the

external scene (e.g., visual background), and the perceiver (e.g., his/her biographical knowledge or processing biases) [(Adams and Kleck, 2003, 2005; Artuso et al., 2021); for a review see Meeren et al. (2005), Wieser and Brosch (2012), Rischer et al. (2020)].

(Adams and Kleck, 2003, 2005; Artuso et al., 2021) for a review see Meeren et al. (2005), Wieser and Brosch (2012), Rischer et al. (2020). Numerous studies have shown that this context effect on facial processing is limited to late latencies (Bradley, 2009; Hajcak et al., 2009, 2010; Diéguez-Risco et al., 2013; Wessing et al., 2013), and that contexts that are conditioned to arousal (particularly threat) increase LPP amplitude (Klein et al., 2015; Xu et al., 2016; Stolz et al., 2019). Of note, these findings disconfirm those of less robust investigations reporting the neural processing of facial expressions and contextual variables as occurring at earlier latencies, involving the vertex positive potential (VPP) and its negative counterpart (N170) (Righart and De Gelder, 2006, 2008; Hietanen and Astikainen, 2013; Diéguez-Risco et al., 2015).

A large body of research has shown that the LPP is also a specific index of memory encoding and storage (Palomba et al., 1997; Schupp et al., 2000), particularly when individuals are confronted with emotionally arousing stimuli (Mecklinger and Pfeifer, 1996; Azizian and Polich, 2007). In line with this, it has been found that meaningful and arousing stimuli elicit larger LPPs and are better retrieved than neutral and relatively less arousing stimuli (Palomba et al., 1997; Dolcos and Cabeza, 2002).

How do emotionally arousing stimuli facilitate memory encoding and storage? Compared to the stimulus valence (i.e., unpleasant–pleasant), which influences relatively early components (100–250 ms), stimulus arousal level (i.e., low–high) influences relatively later (200–1,000 ms) components (Olofsson and Polich, 2007). Moreover, while stimulus valence is indicative of initial selective attention toward salient image content [with unpleasant stimuli producing stronger emotional effects than pleasant stimuli; for a review, see Olofsson et al. (2008)], stimulus arousal leads to an increase in attentional resources that, in turn, facilitates memory encoding and storage (Dolcos and Cabeza, 2002; Schupp et al., 2004a,b).

In the present research, LPP amplitude proved more sensitive in identifying the facial stimuli of patients with different personality pathologies. However, some slight differences merit further discussion. Overall, consistent with the research mentioned above, it may be assumed that LPPs are not merely elicited by patients' faces, *per se*. Instead, they might indicate that, during the EEG task, therapists were engaged in: (a) the attentive processing of affectively meaningful stimuli and “contextual variables” (e.g., Schupp et al., 2006; Bradley, 2009; Hajcak et al., 2009, 2010; Diéguez-Risco et al., 2013; Wessing et al., 2013) related to patients with distinct personality disorders who were previously observed in the videos, and (b)

the memory encoding of patients' (emotionally arousing) faces (Mecklinger and Pfeifer, 1996; Schupp et al., 2000; Azizian and Polich, 2007).

In more detail, the ML models found that amplitudes of specific LPP sub-components were able to accurately discriminate between the facial stimuli of patients with different personality pathologies. Of note, the earliest LPP sub-component (LC1) occurred when therapists were confronted with the facial stimuli of a narcissistic (i.e., Alex) versus a depressive (i.e., April) patient (Figure 5B); the middle LPP sub-component (LC2) occurred when therapists were confronted with the facial stimuli of a narcissistic (i.e., Alex) versus a histrionic/borderline (i.e., Laura) patient (Figure 6B); and the last—and most persistent—LPP sub-component (LC3) occurred when therapists were confronted with the facial stimuli of a depressive (i.e., April) versus a histrionic/borderline (i.e., Laura) patient (Figure 7B).

According to the abovementioned studies on therapist responses to personality-disordered patients, some considerations should be addressed. Narcissistic patients tend to evoke very negative and intense therapist reactions, characterized by hostility, irritation, and contemptuous derogation; histrionic/borderline patients mostly elicit a strong sense of incompetence, inadequacy, confusion, and sexual arousal; whereas depressive patients evoke positive reactions (i.e., nurturing, protection, care). As depicted in Figure 4, the therapist reactions to Alex, Laura, and April were consistent with the empirical literature, suggesting that patients with narcissistic and histrionic/borderline diagnoses show “more arousing” features than depressive patients. Noteworthy, narcissistic and histrionic/borderline diagnoses are among the most challenging clinical syndromes to treat in psychotherapy, especially due to the difficulty of establishing a “good enough” therapist–patient relationship (e.g., Ogrodniczuk and Kealy, 2013; Gabbard, 2014; Caligor et al., 2015; Yeomans et al., 2015).

From this perspective, we may assume that the involvement of the earliest LC1 and LC2 ERP sub-components when therapists were confronted with the facial stimuli of Alex and Laura, respectively (compared with the last and more persistent LC3 sub-component predicting April's face), reflects the particular personality characteristics of these patients. In other words, during the EEG procedure, clinicians' attention processes seem to have been engaged earlier when confronted with (emotionally arousing) visual stimuli connected to Alex and Laura. Conversely, when therapists confronted April's (less arousing) stimuli, their attention and affective meaning attribution processes required later latencies. If future research is able to corroborate these preliminary results, we could argue for the great clinical utility of LPP sub-components in discriminating between patients with distinct affective and interpersonal characteristics.

The present results also showed that, in each tested pair of patients, the LPP was maximally localized in the

hippocampus. This suggests that this cerebral region plays a relevant role in memory processing (Cohen and Eichenbaum, 1995; Eichenbaum, 2001) and the recall of rich and detailed memories (i.e., episodic memories) (e.g., Scoville and Milner, 1957; Addis et al., 2004; Barron et al., 2020). The hippocampus is conceived of as the apex of cortical processing, allowing pattern-separated information to be rapidly bound to support the recall of episodic memories (Felleman and Van Essen, 1991; Mishkin et al., 1998; Addis et al., 2007; Montaldi and Mayes, 2010; Mesulam, 2013). In more detail, the hippocampus receives sensory cortex inputs (including information from all sensory modalities) from the entorhinal, perirhinal, and parahippocampal cortices. On this pathway to the hippocampus, sensory inputs become increasingly elaborated (Pandya and Seltzer, 1982; Van Essen and Maunsell, 1983), and make broad cortico-cortical connections that allow the hippocampal output to be projected back to the neocortex (Witter et al., 1989; Witter, 1993; Squire et al., 2004).

At a physiological level, memory recall implies that, in the presence of an actual retrieval cue, the hippocampal representation of the original experience is promptly reactivated (Wheeler et al., 2000; Rugg and Vilberg, 2013). It has been suggested that, by virtue of reciprocal anatomical connectivity, the hippocampus may coordinate (through excitatory actions) activity in neocortical circuits (Barron et al., 2020), causing the reinstatement of the original neocortical pattern and triggering a re-experience of the original event (Eldridge et al., 2000; Maguire et al., 2001; Moscovitch and McAndrews, 2002; Daselaar et al., 2008; Rugg and Vilberg, 2013).

Of note, beyond sensory information about the external world, interoceptive information regarding the internal and motivational state of the organism from subcortical and brainstem systems is also transmitted to the hippocampus (Strange and Dolan, 2006). It has been proposed that, during memory recall (and particularly during autobiographical memory recall), specific qualities associated with memories (e.g., vivid details, emotionality, personal significance) exert a modulating influence on hippocampal activity and thereby the re-experience of past events (Wheeler et al., 1997; Strange and Dolan, 2006). These assumptions are coherent with the emotional context maintenance and retrieval model (eCMR; Talmi et al., 2019), which extends the original context maintenance and retrieval model (CMR; Polyn et al., 2009) to the emotional domain. In line with this, the eCMR assumes that memory retrieval is not only influenced by the conditions of the retrieval, itself, but also by the qualities of memories with a personal, emotional meaning (Talmi et al., 2019). In summary, the eCMR provides a conceptual explanation of what we need (or are prone) to remember. The hippocampus plays a central role in these processes, given its connectivity with neocortical areas involved in declarative memory (particularly dorsal anterior cingulate regions; Bush et al., 2000; Critchley, 2004) and other cerebral areas (e.g., the

amygdala as well as broad networks underlying attentional and contextual processing) that favor the enhancement of memory for emotional stimuli, during both encoding and retrieval (see Richardson et al., 2004; Dolcos et al., 2012). In light of this evidence, it is plausible to assume that, while observing patients, therapists not only retrieve memories of those patients (and their specific personality disorders), but they also engage attentional and emotional resources to integrate such memories with those of their previous clinical experience with personality disordered patients.

What impact might therapists' previous clinical experience with patients reporting specific personality disorders (e.g., in this case, narcissistic, histrionic/borderline, depressive) have on "real" patients with similar personality diagnoses, such as Alex, Laura, and April? There is mounting empirical evidence to suggest clear connections between memory processes, psychopathology, and psychotherapy (see, e.g., the work of Lane et al., 2015 and all related commentaries), neglecting the ways in which therapists remember or, in other terms, how they use attention and memory processes to serve a healing function [e.g., Gazzaniga, 2008; Boston Change Process Study Group, 2010; for a deeper discussion, see Ekstrom (2014)].

To our knowledge, no prior study has focused on clinicians' memory processes; consequently, based on the extensive neuroscience literature reported above, we could only speculate that, in our pioneering investigation, therapists' meaningful memory recall processes (regarding memories of both the observed patients and previous clinical experiences) affected their ability to distinguish between the facial stimuli of patients with different personality disorders during the EEG procedure. As Addis et al. (2007) underlined, "events in one's past and future are inherently personal and thus should be comprised of autobiographical information" (p. 2). From this premise, and considering the hippocampal role in the reintegration of recollective details in autobiographical memory, Addis et al. suggested that this cerebral structure may also "bind event details for novel future scenarios" (p. 2). This assumption has been repeatedly confirmed by neuroscience research (Okuda et al., 2003; Addis et al., 2007; Rugg and Vilberg, 2013). In fact, the hippocampus comprises part of a common neural network—as well as the medial prefrontal cortex (MPFC) and medial parietal cortex (MPC), extending into the retrosplenial cortex and precuneus (Okuda et al., 2003; Addis et al., 2007; Addis and Schacter, 2008)—that is engaged when both remembering the past and imagining the future.

It has been demonstrated that, in the reconstruction of a past event or the creation of a future event, the hippocampus is engaged early (Addis et al., 2007). Such engagement even precedes that of the PFC, which is known to process self-referential information (Craig et al., 1999; Gusnard et al., 2001; Johnson et al., 2002) such as autobiographical memories (Maguire, 2001; Gilboa, 2004) and imagined future events (Okuda et al., 2003). According to the constructive episodic

simulation hypothesis (Schacter and Addis, 2007), the common neural network for past and future events reflects a reliance on memory to collect details that may be coherent with both past and imagined representations.

Of note, it has also been shown that hippocampal activity is higher when future—as opposed to past—events are imaged (Okuda et al., 2003; Addis et al., 2007). Although both past and future events engage a common neural network aimed at retrieving information from memory, only future scenarios require previously bound information (i.e., memory traces) to be searched, collected, and re-integrated into a novel future event (Cohen et al., 1999; Eichenbaum, 2001). This implies that additional hippocampal resources are needed to allow the details of past events to be successfully bound into a coherent event (Schacter and Addis, 2007).

This relevant role of the hippocampus allows us to extend the clinical implications of our preliminary findings. If processes involving the hippocampus (along with the temporal lobe and other cortical regions) link the past with the present and allow for projection into the future, it may be reasonable to speculate that: (a) therapists' episodic and autobiographical memories (strongly connected to their clinical practice and emotionally significant experiences) exert a modulatory effect on attention processes and the attribution of personal meaning to visual stimuli depicting patients (through complex mechanisms of encoding, storage, and retrieval); and (b) clinicians are able to integrate this information and use semantic knowledge involving personally meaningful representations and symbols to establish accurate patient diagnoses, develop “good enough” therapeutic alliances, and plan effective treatment interventions (Ekstrom, 2014).

The convergence of several empirical and clinical contributions suggests that therapist reactions to narcissistic patients may provoke enactments of judgment, harsh commentary, premature interpretation, criticism, and/or accusatory statements (Gabbard, 2009b; Ronningstam, 2016; cf. also Crisp and Gabbard, 2020); whereas the treatment of histrionic/borderline patients may require firm boundaries to be maintained in the therapeutic relationship, in order to allow for a safe and stable context for patients' self-examination and interpretation of their emotional and relational difficulties (McWilliams and Shedler, 2017). Conversely, when working with depressive patients, the clinical challenge may be the development of an “idealized” patient-therapist relationship in which there is no space for the disappointment and frustration that is inevitable in the therapeutic setting (McWilliams, 2004). Thus, in light of the considerations discussed above, we might assume that clinicians' ability to make accurate clinical decisions (or predictions) in psychotherapy relates to neurophysiological processes that mainly involve the hippocampus.

Despite the reported strengths and implications for future research, the present study has some notable limitations, including the relatively small number of recruited therapists. As regards this point, we wish to underline that the research is

ongoing, and the recruitment of additional cases is underway. A robust enlargement of the therapist sample will allow us to investigate how, in clinicians, therapeutic reactions mediate neural responses during the observation of patient faces. An additional limitation is that distractors (i.e., unfamiliar faces) were not considered in the analyses. We are aware that the inclusion of distractors as control variables would have clarified (and perhaps supported) the above considerations. This is what we intend to verify in the next study.

Bearing in mind these limitations, to the best of our knowledge, no previous study has investigated therapist response patterns to patients with various personality pathologies from a neuroscience perspective. The results of the study, albeit preliminary, hold promise for the field, shedding light on the role played by therapists' memory processes in clinical practice.

Data availability statement

The original contributions presented in this study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Committee of the Department of Dynamic, Clinical Psychology, and Health Studies, Sapienza University of Rome, Italy. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AT conceived the study and wrote clinical issues of the entire manuscript. As first author, she was primarily accountable for all aspects of the work. CT wrote neural issues of the Introduction, Discussion, and Conclusions of the manuscript. AG performed ML analyses, wrote ML results, and provided a substantial contribution to the interpretation of hdEEG data. NC wrote the Introduction section and contributed substantially to the interpretation of the clinical aspects of the study. CC acquired hdEEG data and wrote “hdEEG recordings and data processing” section. CL designed the hdEEG study, monitored hdEEG data acquisition, and wrote “hdEEG recordings and data processing” section. MS-R contributed to the hdEEG study. VL contributed substantially to the conception, design of the work, drafted the manuscript, revised it for intellectual content, and approved its final version to be published. All authors agreed to ensure that the questions that were related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Funding

This study was supported by the Research Project–Prot. No. C26H154HZZB grants from the Sapienza University of Rome, Rome, Italy. Moreover, it received funds from Research Grants to support PDM-oriented research by The Scientific Committee of the Psychodynamic Diagnostic Manual, Second Edition (PDM-2) and the Interdisciplinary Council on Developmental and Learning Disorders.

Acknowledgments

We thank the therapists who participated to this research.

References

- Adams, R. B. Jr., and Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychol. Sci.* 14, 644–647. doi: 10.1046/j.0956-7976.2003.psci.1479.x
- Adams, R. B. Jr., and Kleck, R. E. (2005). Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion* 5, 3–11. doi: 10.1037/1528-3542.5.1.3
- Addis, D. R., Moscovitch, M., Crawley, A. P., and McAndrews, M. P. (2004). Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus* 14, 752–762. doi: 10.1002/hipo.20405
- Addis, D. R., and Schacter, D. L. (2008). Constructive episodic simulation: temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus* 18, 227–237.
- Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45, 1363–1377. doi: 10.1016/j.neuropsychologia.2006.10.016
- Allison, T., Puce, A., Spencer, D. D., and McCarthy, G. (1999). Electrophysiological studies of human face perception: I. potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb. Cortex* 9, 415–430. doi: 10.1093/cercor/9.5.415
- Altavilla, D., Ciacchella, C., Pellicano, G. R., Cecchini, M., Tambelli, R., Kalsi, N., et al. (2021). Neural correlates of sex-related differences in attachment dimensions. *Cogn. Affect. Behav. Neurosci.* 21, 191–211. doi: 10.3758/s13415-020-00859-5
- American Psychiatric Association [APA] (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association Publishing.
- Artuso, C., Palladino, P., and Ricciardelli, P. (2021). Memory updating through aging: different patterns for socially meaningful (and not) stimuli. *Aging Clin. Exp. Res.* 33, 1005–1013. doi: 10.1007/s40520-020-01604-1
- Azizian, A., and Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *J. Cogn. Neurosci.* 19, 2071–2081. doi: 10.1162/jocn.2007.19.12.2071
- Barron, H. C., Auksztulewicz, R., and Friston, K. (2020). Prediction and memory: a predictive coding account. *Prog. Neurobiol.* 192:101821. doi: 10.1016/j.pneurobio.2020.101821
- Bateman, A., and Fonagy, P. (2016). *Mentalization-based Treatment for Personality Disorders: A Practical Guide*. Oxford: Oxford University Press, doi: 10.1093/med:psych/9780199680375.001.0001
- Bateman, A. W., and Fonagy, P. (2006). *Mentalization Based Treatment for Borderline Personality Disorder: A Practical Guide*. New York, NY: Oxford University Press.
- Beck, A. T., Davis, D. D., and Freeman, A. (2004). *Cognitive Therapy of Personality Disorders*, 2nd Edn. New York, NY: Guilford Press.
- Begleiter, H., Porjesz, B., and Garozzo, R. (1979). “Visual evoked potentials and affective ratings of semantic stimuli,” in *Evoked Brain Potentials and Behavior*, ed. H. Begleiter (Boston, MA: Springer), 127–141.
- Berg, J., Lundh, L. G., and Falkenström, F. (2019). Countertransference in Swedish psychotherapists: testing the factor structure of the therapist response questionnaire. *Res. Psychother.* 22:331. doi: 10.4081/ripppo.2019.331
- Betan, E., Heim, A. K., Zittel Conklin, C., and Westen, D. (2005). Countertransference phenomena and personality pathology in clinical practice: an empirical investigation. *Am. J. Psychiatry* 162, 890–898. doi: 10.1176/appi.ajp.162.5.890
- Blagov, P. S., Bi, W., Shedler, J., and Westen, D. (2012). The Shedler-Westen Assessment Procedure (SWAP): evaluating psychometric questions about its reliability, validity, and impact of its fixed score distribution. *Assessment* 19, 370–382. doi: 10.1177/1073191112436667
- Blatt, S. J. (2004). *Experiences of Depression: Theoretical, Clinical, and Research Perspectives*. Washington, DC: American Psychological Association, doi: 10.1037/10749-000
- Blatt, S. J., and Shahar, G. (2004). Psychoanalysis—with whom, for what, and how? comparisons with psychotherapy. *J. Am. Psychoanal. Assoc.* 52, 393–447. doi: 10.1177/00030651040520020401
- Boston Change Process Study Group (2010). *Change in Psychotherapy: A Unifying Paradigm*. New York, NY: W. W. Norton and Company.
- Bourisly, A. K., and Shuaib, A. (2018). Sex differences in electrophysiology: P200 event-related potential evidence. *Transl. Neurosci.* 9, 72–77. doi: 10.1515/tnsci-2018-0013
- Bourke, M. E., and Grenyer, B. F. (2013). Therapists' accounts of psychotherapy process associated with treating patients with borderline personality disorder. *J. Pers. Disord.* 27, 735–745. doi: 10.1521/pedi_2013_27_108
- Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1–11. doi: 10.1111/j.1469-8986.2008.00702.x
- Bush, G., Phan, L., and Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* 4, 215–222. doi: 10.1016/s1364-6613(00)01483-2
- Caligor, E., Levy, K. N., and Yeomans, F. E. (2015). Narcissistic personality disorder: diagnostic and clinical challenges. *Am. J. Psychiatry* 172, 415–422. doi: 10.1176/appi.ajp.2014.14060723
- Carretié, L., Hinojosa, J. A., López-Martín, S., and Tapia, M. (2007). An electrophysiological study on the interaction between emotional content and spatial frequency of visual stimuli. *Neuropsychologia* 45, 1187–1195. doi: 10.1016/j.neuropsychologia.2006.10.013
- Carretié, L., Mercado, F., Hinojosa, J. A., Martín-Loeches, M., and Sotillo, M. (2004). Valence-related vigilance biases in anxiety studied through event-related potentials. *J. Affect. Disord.* 78, 119–130. doi: 10.1016/s0165-0327(02)00242-2

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cecchini, M., Aceto, P., Altavilla, D., Palumbo, L., and Lai, C. (2013). The role of the eyes in processing an intact face and its scrambled image: a dense array ERP and low-resolution electromagnetic tomography (sLORETA) study. *Soc. Neurosci.* 8, 314–325. doi: 10.1080/17470919.2013.797020
- Chertoff, J., Kulish, N., Levinson, N., and Schuker, E. (2020). “Gender issues in transference and countertransference,” in *Female psychology: An annotated psychoanalytic bibliography* (New York, NY: Routledge).
- Clarkin, J. F., Yeomans, F. E., and Kernberg, O. F. (2006). *Psychotherapy for Borderline Personality: Focusing on Object Relations*. Washington, DC: American Psychiatric Association Publishing.
- Clemence, A. J., Perry, J. C., and Plakun, E. M. (2009). Narcissistic and borderline personality disorders in a sample of treatment refractory patients. *Psychiatr. Ann.* 39, 175–184. doi: 10.3928/00485713-20090401-05
- Cohen, N. J., and Eichenbaum, H. (1995). *Memory, Amnesia, and the Hippocampal System*. Cambridge, MA: MIT Press.
- Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., and Nash, C. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus* 9, 83–98. doi: 10.1002/(SICI)1098-1063(1999)9:1<83::AID-HIPO9>3.0.CO;2-7
- Colli, A., Tanzilli, A., Dimaggio, G., and Lingardi, V. (2014). Patient personality and therapist response: an empirical investigation. *Am. J. Psychiatry* 171, 102–108. doi: 10.1176/appi.ajp.2013.13020224
- Conley, E. M., Michalewski, H. J., and Starr, A. (1999). The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. *Clin. Neurophysiol.* 110, 2086–2093. doi: 10.1016/s1388-2457(99)00183-2
- Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., et al. (2018). The racially diverse affective expression (RADATE) face stimulus set. *Psychiatry Res.* 270, 1059–1067. doi: 10.1016/j.psychres.2018.04.066
- Craik, F. I., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., et al. (1999). In search of the self: a positron emission tomography study. *Psychol. Sci.* 10, 26–34. doi: 10.1111/1467-9280.00102
- Crisp, H., and Gabbard, G. O. (2020). Principles of psychodynamic treatment for patients with narcissistic personality disorder. *J. Pers. Disord.* 34(Suppl.), 143–158. doi: 10.1521/pedi.2020.34.supp.143
- Critchley, H. D. (2004). The human cortex responds to an interoceptive challenge. *Proc. Natl. Acad. Sci. U S A.* 101, 6333–6334. doi: 10.1073/pnas.0401510101
- Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., and Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol. Psychol.* 52, 95–111. doi: 10.1016/s0301-0511(99)00044-7
- Dadomo, H., Salvato, G., Lapomarda, G., Ciftci, Z., Messina, I., and Grecucci, A. (2022). Structural features predict sexual trauma and interpersonal problems in borderline personality disorder but not in controls: A multi-voxel pattern analysis. *Front. Hum. Neurosci.* 16:773593. doi: 10.3389/fnhum.2022.773593
- Dahl, H. S., Rössberg, J. I., Bøgwald, K. P., Gabbard, G. O., and Høglend, P. A. (2012). Countertransference feelings in one year of individual therapy: an evaluation of the factor structure in the feeling word Checklist-58. *Psychother. Res.* 22, 12–25. doi: 10.1080/10503307.2011.622312
- Dahl, H.-S. J., Rössberg, J. I., Crits-Christoph, P., Gabbard, G. O., Hersoug, A. G., Perry, J. C., et al. (2014). Long-term effects of analysis of the patient-therapist relationship in the context of patients' personality pathology and therapists' parental feelings. *J. Consult. Clin. Psychol.* 82, 460–471. doi: 10.1037/a0036410
- Daselaar, S. M., Rice, H. J., Greenberg, D. L., Cabeza, R., LaBar, K. S., and Rubin, D. C. (2008). The spatiotemporal dynamics of autobiographical memory: neural correlates of recall, emotional intensity, and reliving. *Cereb. Cortex* 18, 217–229. doi: 10.1093/cercor/bhm048
- Diéguez-Risco, T., Aguado, L., Albert, J., and Hinojosa, J. A. (2013). Faces in context: modulation of expression processing by situational information. *Soc. Neurosci.* 8, 601–620. doi: 10.1080/17470919.2013.834842
- Diéguez-Risco, T., Aguado, L., Albert, J., and Hinojosa, J. A. (2015). Judging emotional congruency: explicit attention to situational context modulates processing of facial expressions of emotion. *Biol. Psychol.* 112, 27–38. doi: 10.1016/j.biopsycho.2015.09.012
- Dolcos, F., and Cabeza, R. (2002). Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. *Cogn. Affect. Behav. Neurosci.* 2, 252–263. doi: 10.3758/cabn.2.3.252
- Dolcos, F., Denkova, E., and Dolcos, S. (2012). Neural correlates of emotional memories: a review of evidence from brain imaging studies. *Psychologia* 55, 80–111. doi: 10.2117/psysoc.2012.80
- Dunning, J. P., and Hajcak, G. (2009). See no evil: directing visual attention within unpleasant images modulates the electrocortical response. *Psychophysiology* 46, 28–33. doi: 10.1111/j.1469-8986.2008.00723.x
- Eichenbaum, H. (2001). The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behav. Brain Res.* 127, 199–207.
- Eimer, M., and Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia* 45, 15–31. doi: 10.1016/j.neuropsychologia.2006.04.022
- Eimer, M., Holmes, A., and McGlone, F. P. (2003). The role of spatial attention in the processing of facial expression: an ERP study of rapid brain responses to six basic emotions. *Cogn. Affect. Behav. Neurosci.* 3, 97–110. doi: 10.3758/cabn.3.2.97
- Ekstrom, S. R. (2014). *Memory and Healing: Neurocognitive and Psychodynamic Perspectives on how Patients and Psychotherapists Remember*. London: Karnac.
- Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., and Engel, S. A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nat. Neurosci.* 3, 1149–1152. doi: 10.1038/80671
- Falkenstein, M., Hoormann, J., and Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychol.* 101, 267–291. doi: 10.1016/s0001-6918(99)00008-6
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1-a
- Foti, D., and Hajcak, G. (2008). Deconstructing reappraisal: descriptions preceding arousing pictures modulate the subsequent neural response. *J. Cogn. Neurosci.* 20, 977–988. doi: 10.1162/jocn.2008.20066
- Foti, D., Hajcak, G., and Dien, J. (2009). Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology* 46, 521–530. doi: 10.1111/j.1469-8986.2009.00796.x
- Frenkel, T. I., and Bar-Haim, Y. (2011). Neural activation during the processing of ambiguous fearful facial expressions: an ERP study in anxious and nonanxious individuals. *Biol. Psychol.* 88, 188–195. doi: 10.1016/j.biopsycho.2011.08.00
- Freud, S. (1910). “The future prospects of psychoanalytic therapy,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, ed. J. Strachey (Richmond, LDN: Hogarth Press), 139–151.
- Freud, S. (1912). “Recommendations to physicians practicing psychoanalysis,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, ed. J. Strachey (Richmond, LDN: Hogarth Press), 111–120.
- Gabbard, G. O. (1995). Countertransference: the emerging common ground. *Int. J. Psychoanal.* 76, 475–485.
- Gabbard, G. O. (2009a). *Textbook of Psychotherapeutic Treatments*. Washington, D.C.: American Psychiatric Association Publishing.
- Gabbard, G. O. (2009b). Transference and countertransference: developments in the treatment of narcissistic personality disorder. *Psychiatr. Ann.* 39, 129–136. doi: 10.3928/00485713-20090301-03
- Gabbard, G. O. (2014). *Gabbard's Treatments of Psychiatric Disorders*, 5th Edn. Arlington, VA: American Psychiatric Publishing.
- Gabbard, G. O., and Westen, D. (2003). Rethinking therapeutic action. *Int. J. Psychoanal.* 84, 823–841. doi: 10.1516/002075703768284605
- Gazzaniga, M. S. (2008). *Human: The Science Behind What Makes Your Brain Unique*. New York, NY: Harper Perennial.
- Ghani, U., Signal, N., Niazi, I. K., and Taylor, D. (2020). ERP based measures of cognitive workload: a review. *Neurosci. Biobehav. Rev.* 118, 18–26. doi: 10.1016/j.neubiorev.2020.07.020
- Gilboa, A. (2004). Autobiographical and episodic memory—one and the same? evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia* 42, 1336–1349. doi: 10.1016/j.neuropsychologia.2004.02.014
- Grecucci, A., Lapomarda, G., Messina, I., Monachesi, B., Sorella, S., and Siugzdaite, R. (2022). Structural features related to affective instability correctly classify the diagnosis of borderline personality disorder. A supervised machine learning approach. *Front. Psychiatry* 13:804440. doi: 10.3389/fpsyt.2022.804440
- Grecucci, A., Sulpizio, S., Tommasello, E., Vespignani, F., and Job, R. (2018). Seeing emotions, reading emotions: behavioral and ERPs evidence of the effect of strategy and of regulation for pictures and words. *PLoS One* 14:e0209461. doi: 10.1371/journal.pone.0209461
- Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U S A.* 98, 4259–4264. doi: 10.1073/pnas.071043098
- Hajcak, G., Dunning, J. P., and Foti, D. (2009). Motivated and controlled attention to emotion: time-course of the late positive potential. *Clin. Neurophysiol.* 120, 505–510. doi: 10.1016/j.clinph.2008.11.028

- Hajcak, G., MacNamara, A., and Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Dev. Neuropsychol.* 35, 129–155. doi: 10.1080/87565640903526504
- Hajcak, G., and Olvet, D. M. (2008). The persistence of attention to emotion: brain potentials during and after picture presentation. *Emotion* 8, 250–255. doi: 10.1037/1528-3542.8.2.250
- Hayes, J. A., Gelso, C. J., Goldberg, S., and Kivlighan, D. M. (2018). Countertransference management and effective psychotherapy: meta-analytic findings. *Psychotherapy* 55, 496–507. doi: 10.1037/pst0000189
- Heimann, P. (1950). On counter-transference. *Int. J. Psychoanal.* 31, 81–84.
- Hennissen, V. C., Meganck, R., Van Nieuwenhove, K., Norman, U. A., Loeys, T., and Desmet, M. (2019). Therapists' responses toward dependent (anaclitic) and self-critical (introjective) depressed outpatients: a multilevel approach. *Psychotherapy* 56, 193–204. doi: 10.1037/pst0000213
- Hietanen, J. K., and Astikainen, P. (2013). N170 response to facial expressions is modulated by the affective congruency between the emotional expression and preceding affective picture. *Biol. Psychol.* 92, 114–124. doi: 10.1016/j.biopsycho.2012.10.005
- Holmes, A., Nielsen, M. K., and Green, S. (2007). Effects of anxiety on the processing of fearful and happy faces: An event-related potential study. *Biol. Psychol.* 77, 159–173. doi: 10.1016/j.biopsycho.2007.10.003
- Hung, Y., Smith, M. L., Bayle, D. J., Mills, T., Cheyne, D., and Taylor, M. J. (2010). Unattended emotional faces elicit early lateralized amygdala-frontal and fusiform activations. *Neuroimage* 50, 727–733. doi: 10.1016/j.neuroimage.2009.12.093
- Jatoui, M. A., Kamel, N., Malik, A. S., and Faye, I. (2014). EEG based brain source localization comparison of sLORETA and eLORETA. *Australas. Phys. Eng. Sci. Med.* 37, 713–721. doi: 10.1007/s13246-014-0308-3
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., and Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain* 125, 1808–1814. doi: 10.1093/brain/awf181
- Kanske, P., and Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Res.* 1148, 138–148. doi: 10.1016/j.brainres.2007.02.044
- Keil, A., Müller, M. M., Gruber, T., Wienbruch, C., Stolarova, M., and Elbert, T. (2001). Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clin. Neurophysiol.* 112, 2057–2068. doi: 10.1016/s1388-2457(01)00654-
- Kernberg, O. (1965). Notes on counter-transference. *J. Am. Psychoanal. Assoc.* 13, 38–56. doi: 10.1177/000306516501300102
- Kernberg, O. F. (1975). *Borderline Conditions and Pathological Narcissism*. Northvale, NJ: Aronson.
- Kernberg, O. F. (2014). An overview of the treatment of severe narcissistic pathology. *Int. J. Psychoanal.* 95, 865–888. doi: 10.1111/1745-8315.12204
- Kissler, J., Assadollahi, R., and Herbert, C. (2006). Emotional and semantic networks in visual word processing: insights from ERP studies. *Prog. Brain Res.* 156, 147–183. doi: 10.1016/S0079-6123(06)56008-X
- Klein, F., Iffland, B., Schindler, S., Wabnitz, P., and Neuner, F. (2015). This person is saying bad things about you: the influence of physically and socially threatening context information on the processing of inherently neutral faces. *Cogn. Affect. Behav. Neurosci.* 15, 736–748. doi: 10.3758/s13415-015-0361-8
- Lai, C., Pellicano, G. R., Ciacchella, C., Guidobaldi, L., Altavilla, D., Cecchini, M., et al. (2020). Neurophysiological correlates of emotional face perception consciousness. *Neuropsychologia* 146:107554. doi: 10.1016/j.neuropsychologia.2020.107554
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated talairach atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.
- Lane, R. D., Ryan, L., Nadel, L., and Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behav. Brain Sci.* 38:e1. doi: 10.1017/S0140525X14000041
- Ligeza, T. S., Maciejczyk, M., Kałamała, P., Szygula, Z., and Wyczesany, M. (2018). Moderate-intensity exercise boosts the N2 neural inhibition marker: a randomized and counterbalanced ERP study with precisely controlled exercise intensity. *Biol. Psychol.* 135, 170–179. doi: 10.1016/j.biopsycho.2018.04.003
- Lingiardi, V., and McWilliams, N. (eds) (2017). *Psychodynamic Diagnostic Manual (PDM-2)*, 2nd Edn. New York, NY: Guilford Press.
- Lingiardi, V., and McWilliams, N. (2015). The psychodynamic diagnostic manual, 2nd ed. (PDM-2). *World Psychiatry* 14, 237–239. doi: 10.1002/wps.20233
- Lingiardi, V., Tanzilli, A., and Colli, A. (2015). Does the severity of psychopathological symptoms mediate the relationship between patient personality and therapist response? *Psychotherapy* 52, 228–237. doi: 10.1037/a0037919
- Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 1, 14–23. doi: 10.1002/widm.8
- Luck, S. (2014). "Event-related potentials," in *APA Handbook of Research Methods in Psychology: Foundations, Planning, Measures, and Psychometrics*, eds H. Cooper, P. M. Camic, D. L. Long, A. Panter, D. Rindskopf, and K. J. Sher (Washington, D.C: American Psychological Association), 1–9.
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Luck, S. J., and Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308. doi: 10.1111/j.1469-8986.1994.tb02218.x
- MacNamara, A., and Hajcak, G. (2009). Anxiety and spatial attention moderate the electrocortical response to aversive pictures. *Neuropsychologia* 47, 2975–2980. doi: 10.1016/j.neuropsychologia.2009.06.026
- MacNamara, A., and Hajcak, G. (2010). Distinct electrocortical and behavioral evidence for increased attention to threat in generalized anxiety disorder. *Depress. Anxiety* 27, 234–243. doi: 10.1002/da.20679
- Maguire, E. A. (2001). Neuroimaging studies of autobiographical event memory. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 1441–1451. doi: 10.1098/rstb.2001.0944
- Maguire, E. A., Frith, C. D., and Cipolotti, L. (2001). Distinct neural systems for the encoding and recognition of topography and faces. *Neuroimage* 13, 743–750. doi: 10.1006/nimg.2000.0712
- Massaro, G., Altavilla, D., Aceto, P., Pellicano, G. R., Lucarelli, G., Luciani, M., et al. (2018). Neurophysiological correlates of collective trauma recall in 2009 L'Aquila earthquake survivors. *J. Trauma. Stress.* 31, 687–697. doi: 10.1002/jts.22334
- McIntyre, S. M., and Schwartz, R. C. (1998). Therapists' differential countertransference reactions toward clients with major depression or borderline personality disorder. *J. Clin. Psychol.* 54, 923–931.
- McWilliams, N. (2004). *Psychoanalytic Psychotherapy: A Practitioner's Guide*. New York, NY: Guilford Press.
- McWilliams, N., and Shedler, J. (2017). "Personality Syndromes—P Axis," in *Psychodynamic Diagnostic Manual, (PDM-2)*, eds V. Lingiardi and N. McWilliams (New York, NY: Guilford Press), 15–74.
- Mecklinger, A., and Pfeifer, E. (1996). Event-related potentials reveal topographical and temporal distinct neuronal activation patterns for spatial and object working memory. *Cogn. Brain Res.* 4, 211–224. doi: 10.1016/s0926-6410(96)00034-1
- Meeren, H. K., van Heijnsbergen, C. C., and de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl. Acad. Sci. U S A.* 102, 16518–16523. doi: 10.1073/pnas.0507650102
- Mesulam, M. M. (2013). Cholinergic circuitry of the human nucleus basalis and its fate in Alzheimer's disease. *J. Comp. Neurol.* 521, 4124–4144. doi: 10.1002/cne.23415
- Mishkin, M., Vargha-Khadem, F., and Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus* 8, 212–216.
- Montaldi, D., and Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus* 20, 1291–1314. doi: 10.1002/hipo.20853
- Moscovitch, D. A., and McAndrews, M. P. (2002). Material-specific deficits in "remembering" in patients with unilateral temporal lobe epilepsy and excisions. *Neuropsychologia* 40, 1335–1342. doi: 10.1016/s0028-3932(01)00213-5
- Mueller, E. M., Hofmann, S. G., Santesso, D. L., Meuret, A. E., Bitran, S., and Pizzagalli, D. A. (2009). Electrophysiological evidence of attentional biases in social anxiety disorder. *Psychol. Med.* 39, 1141–1152. doi: 10.1017/S0033291708004820
- Mühlberger, A., Wieser, M. J., Herrmann, M. J., Weyers, P., Tröger, C., and Pauli, P. (2009). Early cortical processing of natural and artificial emotional faces differs between lower and higher socially anxious persons. *J. Neural Transm.* 116, 735–746. doi: 10.1007/s00702-008-0108-6
- Naumann, E., Bartussek, D., Diedrich, O., and Laufer, M. (1992). Assessing cognitive and affective information processing functions of the brain by means of the late positive complex of the event-related potential. *J. Psychophysiol.* 6, 285–298.
- Nissen-Lie, H. A., Dahl, H. J., and Høglend, P. A. (2022). Patient factors predict therapists' emotional countertransference differently depending on whether

therapists use transference work in psychodynamic therapy. *Psychother. Res.* 32, 3–15. doi: 10.1080/10503307.2020.1762947

Norcross, J. C., and Lambert, M. J. (eds) (2019). *Psychotherapy Relationships that Work: Evidence-based Therapist Contributions*, 3rd Edn. New York, NY: Oxford University Press.

Ogrodniczuk, J. S., and Kealy, D. (2013). “Interpersonal problems of narcissistic patients,” in *Understanding and Treating Pathological Narcissism*, ed. J. S. Ogrodniczuk (Washington, DC: American Psychological Association), 113–127.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., et al. (2003). Thinking of the future and the past: the roles of the frontal pole and the medial temporal lobes. *Neuroimage* 19, 1369–1380. doi: 10.1016/s1053-8119(03)00179-4

Olofsson, J. K., Nordin, S., Sequeira, H., and Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biol. Psychol.* 77, 247–265. doi: 10.1016/j.biopsycho.2007.11.006

Olofsson, J. K., and Polich, J. (2007). Affective visual event-related potentials: arousal, repetition, and time-on-task. *Biol. Psychol.* 75, 101–108. doi: 10.1016/j.biopsycho.2006.12.006

Palomba, D., Angrilli, A., and Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *Int. J. Psychophysiol.* 27, 55–67. doi: 10.1016/s0167-8760(97)00751-4

Pandya, D. N., and Seltzer, B. (1982). Association areas of the cerebral cortex. *Trends Neurosci.* 5, 386–390. doi: 10.1016/0166-2236(82)90219-3

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24(Suppl. D), 5–12.

Pekel, E., and Özmen, E. P. (2020). “Computational intelligence approach for classification of diabetes mellitus using decision tree,” in *Computational Intelligence and Soft Computing Applications in Healthcare Management Science*, eds M. Gul, E. Celik, S. Mete, and F. Serin (Pennsylvania, PA: IGI Global), 87–103. doi: 10.4018/978-1-7998-2581-4.ch005

Perry, J. D., and Perry, J. C. (2004). Conflicts, defenses and the stability of narcissistic personality features. *Psychiatry* 27, 310–330. doi: 10.1521/psyc.67.4.310.56570

Peterson, N. N., Schroeder, C. E., and Arezzo, J. C. (1995). Neural generators of early cortical somatosensory evoked potentials in the awake monkey. *Electroencephalogr. Clin. Neurophysiol.* 96, 248–260. doi: 10.1016/0168-5597(95)00006-e

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127

Pires, L., Leitão, J., Guerrini, C., and Simões, M. R. (2014). Event-related brain potentials in the study of inhibition: cognitive control, source localization and age-related modulations. *Neuropsychol. Rev.* 24, 461–490. doi: 10.1007/s11065-014-9275-4

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Polyn, S. M., Norman, K. A., and Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychol. Rev.* 116:129. doi: 10.1037/a0014420

Richardson, M. P., Strange, B. A., and Dolan, R. J. (2004). Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nat. Neurosci.* 7, 278–285. doi: 10.1038/nn1190

Righart, R., and De Gelder, B. (2006). Context influences early perceptual analysis of faces—an electrophysiological study. *Cereb. Cortex* 16, 1249–1257. doi: 10.1093/cercor/bhj066

Righart, R., and De Gelder, B. (2008). Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Soc. Cogn. Affect. Neurosci.* 3, 270–278. doi: 10.1093/scan/nsn021

Rischer, K. M., Savallampi, M., Akwathage, A., Salinas Thunell, N., Lindersson, C., and MacGregor, O. (2020). In context: emotional intent and temporal immediacy of contextual descriptions modulate affective ERP components to facial expressions. *Soc. Cogn. Affect. Neurosci.* 15, 551–560. doi: 10.1093/scan/nsaa071

Roche, R. A., Dockree, P. M., Garavan, H., Foxe, J. J., Robertson, I. H., and O'Mara, S. M. (2004). EEG alpha power changes reflect response inhibition deficits after traumatic brain injury (TBI) in humans. *Neurosci. Lett.* 362, 1–5. doi: 10.1016/j.neulet.2003.11.064

Ronningstam, E. (2012). Alliance building and narcissistic personality disorder. *J. Clin. Psychol.* 68, 943–953. doi: 10.1002/jclp.21898

Ronningstam, E. (2016). Pathological narcissism and narcissistic personality disorder: recent research and clinical implications. *Curr. Behav. Neurosci. Rep.* 3, 34–42.

Rossberg, J. I., Karterud, S., Pedersen, G., and Friis, S. (2007). An empirical study of countertransference reactions toward patients with personality disorders. *Compr. Psychiatry* 48, 225–230. doi: 10.1016/j.comppsy.2007.02.002

Rossberg, J. I., Karterud, S., Pedersen, G., and Friis, S. (2008). Specific personality traits evoke different countertransference reactions: an empirical study. *J. Nerv. Ment. Dis.* 196, 702–708.

Rugg, M. D., and Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. *Curr. Opin. Neurobiol.* 23, 255–260. doi: 10.1016/j.conb.2012.11.005

Sabatini, D., Lang, P. J., Keil, A., and Bradley, M. M. (2007). Emotional perception: correlation of functional MRI and event-related potentials. *Cereb. Cortex* 17, 1085–1091. doi: 10.1093/cercor/bhl017

Schacter, D. L., and Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 362, 773–786. doi: 10.1098/rstb.2007.2087

Schupp, H. T., Cuthbert, B., Bradley, M., Hillman, C., Hamm, A., and Lang, P. (2004a). Brain processes in emotional perception: motivated attention. *Cogn. Emot.* 18, 593–611. doi: 10.1080/02699930341000239

Schupp, H. T., Junghöfer, M., Weike, A. I., and Hamm, A. O. (2004b). The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* 41, 441–449. doi: 10.1111/j.1469-8986.2004.00174.x

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., and Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology* 37, 257–261. doi: 10.1111/1469-8986.3720257

Schupp, H. T., Flaisch, T., Stockburger, J., and Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Prog. Brain Res.* 156, 31–51. doi: 10.1016/S0079-6123(06)56002-9

Schwartz, R. C., Smith, S. D., and Chopko, B. (2007). Psychotherapists' countertransference reactions toward clients with antisocial personality disorder and schizophrenia: an empirical test of theory. *Am. J. Psychother.* 61, 375–393. doi: 10.1176/appi.psychotherapy.2007.61.4.375

Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21. doi: 10.1136/jnnp.20.1.11

Shedler, J., and Westen, D. (2004). Dimensions of personality pathology: an alternative to the five-factor model. *Am. J. Psychiatry* 161, 1743–1754. doi: 10.1176/ajp.161.10.1743

Shedler, J., Westen, D., and Lingardi, V. (2014). *The Evaluation of Personality with the SWAP-200*. Milan: Raffaello Cortina.

Sokhadze, E. M., Casanova, M. F., Casanova, E. L., Lamina, E., Kelly, D. P., and Khachidze, I. (2017). Event-related potentials (ERP) in cognitive neuroscience research and applications. *Neuroregulation* 4, 14–14. doi: 10.15540/nr.4.1.14

Squire, L. R., Stark, C. E., and Clark, R. E. (2004). The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306. doi: 10.1146/annurev.neuro.27.070203.144130

Stefana, A., Bulgari, V., Youngstrom, E. A., Dakanalis, A., Bordin, C., and Hopwood, C. J. (2020). Patient personality and psychotherapist reactions in individual psychotherapy setting: a systematic review. *Clin. Psychol. Psychother.* 27, 697–713. doi: 10.1002/cpp.2455

Stolz, C., Endres, D., and Mueller, E. M. (2019). Threat-conditioned contexts modulate the late positive potential to faces—a mobile EEG/virtual reality study. *Psychophysiology* 56:e13308. doi: 10.1111/psyp.13308

Strange, B. A., and Dolan, R. J. (2006). Anterior medial temporal lobe in human cognition: memory for fear and the unexpected. *Cogn. Neuropsychiatry* 11, 198–218. doi: 10.1080/13546800500305096

Streiner, D. L. (2003). Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *J. Pers. Assess.* 80, 217–222. doi: 10.1207/S15327752JPA8003_01

Sulpizio, S., Toti, M., Del Maschio, N., Costa, A., Fedeli, D., Job, R., et al. (2019). Are you really cursing? neural processing of taboo words in native and foreign language. *Brain Lang.* 194, 84–92. doi: 10.1016/j.bandl.2019.05.003

Sur, S., and Sinha, V. K. (2009). Event-related potential: an overview. *Ind. Psychiatry J.* 18, 70–73. doi: 10.4103/0972-6748.57865

Talmi, D., Lohnas, L. J., and Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychol. Rev.* 126, 455–485. doi: 10.1037/rev0000132

- Tanner, D., Morgan-Short, K., and Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology* 52, 997–1009. doi: 10.1111/psyp.12437
- Tanzilli, A., Colli, A., Del Corno, F., and Lingardi, V. (2016). Factor structure, reliability, and validity of the therapist response questionnaire. *Pers. Disord. Theory Res. Treat.* 7, 147–158. doi: 10.1037/per0000146
- Tanzilli, A., and Gualco, I. (2020). Clinician emotional responses and therapeutic alliance when treating adolescent patients with narcissistic personality disorder subtypes: A clinically meaningful empirical investigation. *J. Pers. Disord.* 34, 42–62. doi: 10.1521/pedi.2020.34.supp.42
- Tanzilli, A., Gualco, I., Baiocco, R., and Lingardi, V. (2020). Clinician reactions when working with adolescent patients: the therapist response questionnaire for adolescents. *J. Pers. Assess.* 102, 616–627. doi: 10.1080/00223891.2019.1674318
- Tanzilli, A., Lingardi, V., and Hilsenroth, M. (2018). Patient SWAP-200 personality dimensions and FFM traits: do they predict therapist responses? *Pers. Disord. Theory Res. Treat.* 9, 250–262. doi: 10.1037/per0000260
- Tanzilli, A., Muzi, L., Ronningstam, E., and Lingardi, V. (2017). Countertransference when working with narcissistic personality disorder: an empirical investigation. *Psychotherapy* 54, 184–194. doi: 10.1037/pst0000111
- Van Essen, D. C., and Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci.* 6, 370–375.
- Vogel, E. K., and Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology* 37, 190–203. doi: 10.1111/1469-8986.3720190
- Vuilleumier, P., and Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: evidence from functional neuroimaging. *Neuropsychologia* 45, 174–194.
- Wampold, B. E. (2015). How important are the common factors in psychotherapy? an update. *World Psychiatry* 14, 270–277. doi: 10.1002/wps.20238
- Weinberg, A., and Hajcak, G. (2010). Beyond good and evil: the time-course of neural activity elicited by specific picture content. *Emotion* 10, 767–782. doi: 10.1037/a0020242
- Weinberg, A., and Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *J. Cogn. Neurosci.* 23, 2994–3007. doi: 10.1162/jocn.2011.21630
- Wessing, I., Rehbein, M. A., Postert, C., Färniss, T., and Junghöfer, M. (2013). The neural basis of cognitive change: reappraisal of emotional faces modulates neural source activity in a frontoparietal attention network. *Neuroimage* 81, 15–25. doi: 10.1016/j.neuroimage.2013.04.117
- Westen, D., and Muderrisoglu, S. (2006). Clinical assessment of pathological personality traits. *Am. J. Psychiatry* 163, 1285–1287. doi: 10.1176/ajp.2006.163.7.1285
- Westen, D., and Shedler, J. (1999a). Revising and assessing axis II, Part I: developing a clinically and empirically valid assessment method. *Am J Psychiatry* 156, 258–272. doi: 10.1176/ajp.156.2.258
- Westen, D., and Shedler, J. (1999b). Revising and assessing axis II, Part II: toward an empirically based and clinically useful classification of personality disorders. *Am. J. Psychiatry* 156, 273–285. doi: 10.1176/ajp.156.2.273
- Wheeler, M. A., Stuss, D. T., and Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol. Bull.* 121, 331–354. doi: 10.1037/0033-2909.121.3.331
- Wheeler, M. E., Petersen, S. E., and Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *PNAS* 97, 11125–11129. doi: 10.1073/pnas.97.20.11125
- Wieser, M. J., and Brosch, T. (2012). Faces in context: a review and systematization of contextual influences on affective face processing. *Front. Psychol.* 3:471. doi: 10.3389/fpsyg.2012.00471
- Witter, M. P. (1993). Organization of the entorhinal–hippocampal system: a review of current anatomical data. *Hippocampus* 3, 33–44.
- Witter, M. P., Groenewegen, H. J., Lopes, da Silva, F. H., and Lohman, A. H. (1989). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Prog. Neurobiol.* 33, 161–253. doi: 10.1016/0301-0082(89)90009-9
- Xu, M., Li, Z., Diao, L., Fan, L., and Yang, D. (2016). Contextual valence and sociality jointly influence the early and later stages of neutral face processing. *Front. Psychol.* 7:1258. doi: 10.3389/fpsyg.2016.01258
- Yeomans, F. E., Clarkin, J. F., and Kernberg, O. F. (2015). *Transference Focused Psychotherapy for Borderline Personality Disorder: A Clinical Guide*. Washington, DC: American Psychiatric Publishing.
- Zittel Conklin, C., and Westen, D. (2003). *The Therapist Response Questionnaire (TRQ)*. Atlanta, GA: Departments of Psychology and Psychiatry and Behavioral Sciences, Emory University.



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Yorgos Dimitriadis,
Université de Paris, France
Maria Cristina Patru,
International Mental Health
SA, Switzerland

*CORRESPONDENCE

Jessica Tran The
jessica.tran-the@unicaen.fr

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 30 May 2022

ACCEPTED 21 November 2022

PUBLISHED 14 December 2022

CITATION

Tran The J, Ansermet J-P,
Magistretti PJ and Ansermet F (2022)
Hyperactivity of the default mode
network in schizophrenia and free
energy: A dialogue between Freudian
theory of psychosis and neuroscience.
Front. Hum. Neurosci. 16:956831.
doi: 10.3389/fnhum.2022.956831

COPYRIGHT

© 2022 Tran The, Ansermet,
Magistretti and Ansermet. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Hyperactivity of the default mode network in schizophrenia and free energy: A dialogue between Freudian theory of psychosis and neuroscience

Jessica Tran The^{1,2,3,4,5,6*}, Jean-Philippe Ansermet⁷,
Pierre J. Magistretti^{5,8,9} and Francois Ansermet^{5,10}

¹INSERM U1077 Neuropsychologie et Imagerie de la Mémoire Humaine, Caen, France, ²Centre Hospitalier Universitaire de Caen, Caen, France, ³Université de Caen Normandie, Caen, France, ⁴Ecole Pratique des Hautes Etudes, Université Paris Sciences et Lettres, Paris, France, ⁵Agalma Foundation Geneva, Geneva, Switzerland, ⁶Cyceron, Caen, France, ⁷École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, ⁸Division of Biological and Environmental Sciences and Engineering (BESE), King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, ⁹Brain Mind Institute, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland, ¹⁰Département de Psychiatrie, Faculté de Médecine, Université de Genève, Geneva, Switzerland

The economic conceptualization of Freudian metapsychology, based on an energetics model of the psyche's workings, offers remarkable commonalities with some recent discoveries in neuroscience, notably in the field of neuroenergetics. The pattern of cerebral activity at resting state and the identification of a *default mode network* (DMN), a network of areas whose activity is detectable at baseline conditions by neuroimaging techniques, offers a promising field of research in the dialogue between psychoanalysis and neuroscience. In this article we study one significant clinical application of this interdisciplinary dialogue by looking at the role of the DMN in the psychopathology of schizophrenia. Anomalies in the functioning of the DMN have been observed in schizophrenia. Studies have evidenced the existence of hyperactivity in this network in schizophrenia patients, particularly among those for whom a positive symptomatology is dominant. These data are particularly interesting when considered from the perspective of the psychoanalytic understanding of the positive symptoms of psychosis, most notably the Freudian hypothesis of delusions as an "attempt at recovery." Combining the data from research in neuroimaging of schizophrenia patients with the Freudian hypothesis, we propose considering the hyperactivity of the DMN as a consequence of a process of massive reassociation of traces occurring in schizophrenia. This is a process that may constitute an attempt at minimizing the excess of free energy present in psychosis. Modern models of active inference and the free energy principle (FEP) may shed some light on these processes.

KEYWORDS

default network, schizophrenia, trace reassociation, psychoanalysis, free energy

Introduction

The economic conceptualization of Freudian metapsychology, which is based on an energetics model of the psyche's workings, offers remarkable commonalities with some recent advances in neuroscience, notably discoveries in the field of neuroenergetics (Friston et al., 2006; Tran The et al., 2018; Cieri and Esposito, 2019). As early as 1895, in his *Project for a Scientific Psychology* (1895a), Freud's intention had been to base his understanding of normal and pathological psychic processes on a description of brain function. At that time, he had postulated the existence of a fundamental tendency of the nervous system to reduce all quantities of neuronal excitation: "the principle of neuronal inertia." This speculative proposal on brain function rested on a concept of "quantity" of neuronal energy or intracerebral excitation a theory influenced both by the physicalism of the Helmholtz school (Helmholtz, 1882), and by Oswald's energetics model. Oswald stated that: "...psychological phenomena can be construed as energetic phenomena, and interpreted as such just as well as any other phenomena" (Oswald, 1891). This desire to apply the overarching principles of thermodynamic physics to living organisms, and more specifically to the nervous system, supports the view that the father of psychoanalysis was following in the tradition of the Helmholtz School and its director, Brücke.

Yet, Freud will distance himself from this ambition to explain the psychological processes by reference to the physiology of the nervous system. He did not proceed with the publication of the *Project* manuscript, which remained unfinished, because of the limitations of biological knowledge at that time:

"[I am] not at all inclined to leave the psychology hanging in the air without an organic basis. But apart from this conviction I do not know how to go on, neither theoretically nor therapeutically, and therefore must behave as if only the psychological were under consideration." (Freud, 1898, p. 326)

Thus, from 1897, with *The Interpretation of Dreams* (Freud, 1900), Freud abandoned his references to anatomy and adopted an exclusively psychological vocabulary. It is at that point that he introduced the fundamental concepts of his metapsychology or theory of the unconscious psychic processes. He did this, however, without completely forgoing his energetics concept of the psychic processes, since the quantity of psychic energy becomes the concept of *libido*.

"I refer to the concept that in mental functions something is to be distinguished – a quota of effect or sum of excitation – which possesses all the characteristics of a quantity (though we have no means of measuring it), which is capable of increase, diminution, displacement and

discharge, and which is spread over the memory-traces of ideas somewhat as an electric charge is spread over the surface of a body." (Freud, 1894, p. 60)

Although Freud had renounced his hope of being able to measure that quantity of energy and link it to processes in the brain (Rapaport, 1958), the psychoanalytic concept of the psychic processes in terms of energy shows some surprising commonalities with the field of neuroenergetics. This new field of research has been opened up by the technical advances made in neuroscience, notably in the area of neuroimaging. In 1920, Freud himself had imagined the feasibility of such an interdisciplinary dialogue:

"Biology is truly a land of unlimited possibilities. We may expect it to give us the most surprising information and we cannot guess what answers it will return in a few dozen years to the questions we have put to it." (Freud, 1920, p. 60)

If psychoanalysis and neuroscience remain two fields with their own epistemological foundations, today these two fields do seem to have found some common ground in an energetics concept of the living organism (Ansermet and Magistretti, 2010). From a strictly biological perspective energetics has become one of the leading ideas. Recent neuroscientific research has brought to light important data relating to what could be qualified as an intrinsic activity belonging to living organisms. In particular, the developments of recent techniques in neuroimaging such as positive emission tomography (PET) or functional magnetic resonance imaging (fMRI), has offered scientists the possibility to observe regions of the brain activated during the performing of specific tasks—be they motor, sensorial or cognitive tasks.

Since the 2000's, some scientists have started to take an interest in the background brain activity. This is activity that takes place during states of rest when no task or specific behavior is taking place. In particular, Marcus Raichle has demonstrated that, at rest (i.e., not performing any particular activity), some areas of the brain are more active than during the performing of a specific task (Raichle and Gusnard, 2002). There exists therefore an activity intrinsic to the brain, a background activity, one that is particularly costly in energy. This cerebral network, active during *resting state* and called *default mode network* (DMN), presents a promising field of research for the dialogue between psychoanalysis and neuroscience around a concept of the living organism centered on energetics. And we can try to throw light on the Freudian energetics approach to psychopathological processes with these recent neuroscientific data.

In the *Project* (1895a), Freud's energetics concept was initially principally motivated by his clinical observations of neurotic patients. He stated that his concept of quantity "is derived directly from pathological clinical observation" and that in patients with neuroses "the quantitative characteristic emerges more plainly than in the normal." (Freud, 1895, p. 295).

In his *Studies on Hysteria* (1895d), Freud put forward a theory of the psychopathology of hysteria as being a consequence of the increase in the quantity of energy (psychic excitation or *quantum of affect*). This quantity of excitation was then discharged by somatic means, through the expression of symptoms of conversion (Freud and Breuer, 1895). In a more general way, as his texts on the neuro-psychoses of defense demonstrate (Freud, 1894), right from the start Freud understood the different psychic pathologies as being linked to pathological alterations in quantities of psychic energy (later termed libido). These alterations could be linked to acquired factors (for example subsequent to trauma) (Freud and Breuer, 1895), as well as hereditary factors, insofar as in certain subjects heredity “Facilitates and increases the pathological affect” (Freud, 1896b, p. 162) in the same way as “...a multiplier in an electric circuit, which exaggerates the visible deviation of the needle,” (Freud, 1896a, p. 147).

Subsequently, during his study of the Schreber case, Freud extended this hypothesis for a libidinal etiology of psychic pathologies to the treatment of psychosis. At that moment he proposed an innovative theory of this illness, based on a diachronic perspective, by arguing that psychosis stemmed from a libidinal process in two phases. The initial process of psychosis consisted in massive withdrawal of the libido (or quantity of psychic energy) into the ego and the body itself (Freud, 1914). This was the first phase of the psychotic process, dominated by significant alteration in the perception of the body. The *positive* symptoms, and importantly the patients’ delusional constructions, belong to the second phase of psychosis, and came as an attempt to reinvest the quantity of energy externally, on objects and delusional representations. Delusion was thus, in Freud’s view, an “attempt at recovery” (Freud, 1911).

This Freudian perspective of a libidinal etiology for psychosis, which fits into the framework of a concept of pathologies of the psyche based on energetics (and the “economics” view of the metapsychology), can be integrated into neuroscientific discoveries about DMN, and into recent integrated models of brain function. In the *Project*, Freud introduced energy quantities based on language of Helmholtzian thermodynamics (Tran The et al., 2018). Nowadays, a new conceptual framework has emerged from Bayesian statistics and information theory (Parr et al., 2022). Thus, Friston and coworkers proposed a dialog between cognitive science, neuroscience and psychoanalysis, based on the notion of Active Inference which is rendered quantitative thanks to the principle of minimization of a free energy (Carhart-Harris and Friston, 2019).

Karl Friston and others have proposed introducing a new model for a dialogue between psychoanalysis and neuroenergetics based on the FEP. The FEP proposes that living organisms, as organized systems in a state of equilibrium with their environment, need to minimize their free energy, which is to say they need to resist the natural tendency to disorder (Friston, 2009; Ramstead et al., 2018). As has been suggested by

Mark Solms, this minimizing of the free energy is similar to a homeostatic understanding of the regulation of the physiological states, which would be one of the most basic functions of consciousness-based somatic and affective states (Solms, 2013; Solms and Friston, 2018). In Solms’ words, “...the long-sought mechanism of consciousness is to be found in an extended form of homeostasis” (Solms, 2019). By integrating the idea of free energy into a Bayesian approach to cognition, where the brain is understood as a machine for inference (Helmholtz, 1866; Gregory, 1968; Dayan et al., 1995), this model proposes that the brain uses internal hierarchical models to predict sensorial inputs. The hypothesis is that neuronal activity and synaptic connections try to minimize the ensuing prediction-error resulting from the inference processes (or try to minimize the free energy) (Carhart-Harris and Friston, 2010; Parr et al., 2022). But the FEP does not only constitute an epistemological focus in the interdisciplinary dialogue. It also opens onto important clinical applications in the field of psychopathology (Rabeyron, 2022).

From this standpoint, we will try to open a dialogue between Freud’s view of delusion as an “attempt at recovery” and recent data and models coming from the neurosciences, which have been obtained through to the techniques of neuroimaging. The data attest to the existence, in some psychiatric illnesses, of anomalies in the neuroenergetic activity of the brain. In this present article, we focus particularly on the literature that evidences the existence of hyperactivity of the DMN in schizophrenia, that is to say: an abnormally increased activity in energetic terms observed in the cerebral activity at rest, in patients with schizophrenia. In this article we propose using an interdisciplinary approach to shed light on these data, through the prism of a dialogue between psychoanalysis and neuroscience. To do this we will make reference to the Freudian hypothesis of delusion as an “attempt at recovery,” and integrate these elements to the FEP model.

To this end, we will first give an overview of the discovery of the DMN and the hypotheses relative to the DMN’s possible functions. We will identify some common ground between the DMN and the binding function of the ego in Freudian theory, into the framework of the FEP. Then we will review the anomalies in this network observed in schizophrenia. Next, having reviewed the Freudian concept of the positive symptoms of psychosis—notably his hypothesis of delusion as an attempt at recovery— we will bring together the data on the hyperactivity of the DMN in patients with schizophrenia and these Freudian hypotheses. Through this reflection, we propose interpreting the hyperactivity in the DMN as consequent to a massive process of reassociation of traces occurring in schizophrenia. The FEP perspective might be invoked to clarify the dynamics of this particular process. In this light, delusion could constitute an attempt at binding the free energy through the reassociation of traces, so as to minimize entropy and try to re-establish an allostatic equilibrium within the psyche.

The discovery of cerebral activity during resting state and the default mode network

Initial research in neuroenergetics, undertaken by Louis Sokoloff and Seymour Kety in the 1950's, already revealed that the cost of internal functional activity represented a significant part of the total energy expenditure of cerebral activity (Sokoloff et al., 1955). While the average adult brain represents ~2% of the total body weight, it nevertheless uses 20–25 % of the organism's total energy consumption—and that, even during resting state (Sokoloff et al., 1955). The performing of tasks therefore, has only a minimal incidence on the massive energy consumption of the brain, producing only a rise of barely 5–10 % of the total consumption (Fox and Raichle, 2007). Marcus Raichle, in the 1990's, began to envisage this baseline activity as a promising field for research (Raichle, 2006), and a source of new data for the understanding of certain psychopathological processes (Fox and Raichle, 2007).

The development of different imaging techniques, such as PET and fMRI made it possible to identify a network of active areas when cognitive tasks were not being performed: the medial prefrontal cortex, posterior cingulate cortex, the inferior parietal lobules and the hippocampal regions (Raichle et al., 2001; Andrews-Hanna et al., 2014). The work done by Raichle, on the basis of the data furnished by the study of neuroimaging, led to the description and theorisation of the DMN. A default mode that attests to the existence of neural and metabolic activity that is intrinsic to the brain, and which is not linked either to a particular environmental stimulus, nor to the performing of a specific motor or cognitive task (Buckner et al., 2008).

In the state of consciousness termed random episodic silent thinking (REST)—during which the individual processes corporeal information (somatosensory and vegetative), and during which the individual also free associates ideas, emotions, and mental images connected to the past or to future planning—the attentional system is not solicited, and it is the DMN that is active. This is the spontaneous brain activity generated in the absence of any cognitive task (Raichle, 2001; Buckner et al., 2008; Cieri and Esposito, 2018, 2019).

Raichle proposed defining this default mode, or baseline state, as the brain activity that goes with a subjective state of rest, when the subject presents a behavior characterized by calmness, eyes closed (or open but with no specific focus), in a state of consciousness filled with thoughts that are independent of external stimuli, a state that might be described as “daydreaming” (Raichle, 2010). Based on the data from functional imaging, the brain is in fact never really at rest, insofar as the consumption of energy linked to its intrinsic activity, independent of any action or perception of external stimuli, is extremely high. This baseline activity can be observed even under general anesthesia, or in the first stages of sleep.

Furthermore, one of the major discoveries made by Raichle and his team was to bring to light a diminution of the default mode activity during the performing of tasks that require directed attention. That is to say that, certain brain regions active at rest deactivate during the performing of a conscious goal-directed task (Shulman et al., 1997; Raichle, 2010; Passow et al., 2015).

This reciprocal deactivating of the DMN and the networks linked to the attention system, such as the dorsal attention network (DAN), which appears to be anticorrelated (Fransson, 2005), involves an opposition between an event-related brain activity and spontaneous brain activity (evoked/spontaneous). The latching mechanism between the two networks may demonstrate a change in attention that switches from an internal self-referential state to an outwardly focused state (Raichle, 2001), with a shift of attention between the internal and the external. The attention system, converse to the DMN, is activated when cognition is directed outwards and the opposite is true for the DMN (Buckner et al., 2008). However, recent research indicates that the attention system is not a unified one. Besides the DAN, the salience network is also active during the performing of goal-directed tasks (Corbetta and Shulman, 2002; Fox et al., 2006; Carhart-Harris and Friston, 2010; Esposito et al., 2018).

The mechanisms that contribute to the deactivation of the DMN during the execution of a task that is demanding at a cognitive level, and involves focused attention on the environment, are still poorly understood. There are however some studies that have brought to light a significant co-activation of a third network during the “switch” between the activity of the DMN and that of the attention system (Sridharan et al., 2008). The salience network (a network comprising the right insular cortex and the anterior cingulate cortex) is responsible for switching between the default mode network and the central executive network (Goulden et al., 2014). Scientists have therefore concluded that the insular cortex plays a key role in this switch over between the internal cerebral activity and activity that is directed toward the environment; and is probably linked to the importance of the insula in distinguishing between endogenous processes and exteroceptive perceptions (Craig, 2002; Tran The et al., 2020).

Hypotheses relating to the functions of the default mode

To better understand the mechanisms of the DMN, Raichle initially based his thinking on what was already known regarding the behavioral and cognitive functions associated with the main anatomical subdivisions of the baseline state network, that is: the ventromedial prefrontal cortex, the dorsomedial prefrontal cortex, and the posterior cingulate cortex, the adjacent precuneus, the lateral parietal cortex (Raichle, 2015).

The ventromedial prefrontal cortex in particular is involved in the processing of sensory information from the environment, coming from the orbitofrontal cortex, and then transmitted to structures such as the hippocampus and the amygdala (Raichle, 2015). Data from patients having suffered brain damage have highlighted the importance of this area in the control of social behavior, mood, and motivation (Damasio et al., 1994). Furthermore, studies of imaging performed on healthy individuals have shown that the emotional state (as well as the anxiety levels) of the subject has a direct effect on the levels of activity in the ventromedial prefrontal cortex within the baseline state's network (Park and Moghaddam, 2017). An increase in the activity of the areas of the ventromedial prefrontal cortex can also be observed during disturbances in the body's homeostasis, for example during a state of hypoglycaemia (Teves et al., 2004). It can be supposed therefore that the baseline state, in that its network involves the activation of the ventromedial prefrontal cortex, may be involved in the processing of emotions and disruption to homeostasis. The dorsomedial prefrontal cortex, although it is adjacent to the ventromedial prefrontal cortex, is involved more specifically in the perception of the self, and in the processing of emotions that accompany that perception of the self (Raichle, 2015).

The posterior cingulate cortex and the medial precuneus make up another area, notably associated with lateral parietal elements of the DMN, that is particularly active during the baseline state. A study made by Vincent et al. has brought to light that these specific areas of the parietal cortex are involved in episodic memory. Studies of functional imaging attest to their systematic activation when a memorizing task has successfully been performed following learning (Vincent et al., 2006). This study has also contributed to highlighting the existence of a neuronal network linking these parietal areas to the hippocampus, all of these networks being activated when the subject is at rest in the absence of a task or explicit mnemonic stimulation. This activity is particularly increased if the resting state is consecutive to activities involving memory or the recall of elements previously memorized (Shannon et al., 2013).

More generally, the study of these data on the functioning of the different anatomical areas belonging to the DMN indicate that it are linked to thought processes such as daydreaming (Buckner et al., 2008), self-referential mental activity (for example attention focused on emotional or physical states), the processing of autobiographical memories, or thoughts about the future (Doucet et al., 2011; Dor-Ziderman et al., 2013), and theory of mind (Carhart-Harris and Friston, 2019). This has led scientist to hypothesis that the DMN could be the neurobiological seat of the 'self' (Andrews-Hanna, 2012; Cieri and Esposito, 2019).

In so far as that the activity of the default mode seems to correlate strongly with the recollection of autobiographical memories (Vincent et al., 2006; Buckner et al., 2008; Doucet et al., 2011), this cerebral network could constitute one of the

biological foundations of what Damasio has defined as our awareness of an autobiographical self. This autobiographical self is underpinned by old memories, the stability of which guarantees the grounds for our sense of identity regardless of the continuous changes to which we are subject (Damasio, 2000). From the perspective of *embodied cognition*, the sense of our autobiographical self and identity is also intimately dependent on certain representations linked to our physical cartography, to the mental image of our own body, within certain areas of the brain. The neuronal mechanisms that generate this representation can be seen as the foundation for a kind of stability of the organism—which Damasio terms the “proto-self”—that is at the root of our sense of self. It is on the basis of this stable representation of the body—an anchor from the point of view of an invariable reference through time—that the sense of an autobiographical self will subsequently develop. This is an autobiographical self from which stems our sense of identity regardless of the continuous changes that shape us and that:

...might give the brain a natural means to generate the singular and stable reference we call self [...] biologically speaking, grounded on a collection of nonconscious neural patterns standing for the part of the organism we call the body proper. (Damasio, 2000, pp. 134-135)

Within the DMN certain areas may be involved in the integration of external stimuli, such as interoceptive physical sensation, as well as information coming from the environment, for example visual and auditory perceptions (Phan et al., 2002). In particular, the interaction between the DMN and the insula may contribute to this integration between the external and interoceptive sensations (Molnar-Szakacs and Uddin, 2013; Cieri and Esposito, 2019). Research on the anterior insula undertaken by Craig has made it possible to characterize the neurobiological foundations of this “interoceptive” sense, which furnishes the organism with a representation in real time of the state of its various organs. Indeed, even in the absence of external stimuli from the environment, the brain constantly processes endogenous perceptions coming from the viscera, the muscles, and the joints, as well as the different concentrations of hormones in the blood stream, which are constantly being detected by the nervous system. Information coming from the body is initially processed in the medial and posterior areas of the insula then, in a second phase, integrated within the anterior insula (Craig, 2002). The integration between the perception of exteroceptive experiences and the representations coming from the interoceptive endogenous perceptions, in so far as that it is a process staggered in time in relation to the initial experience (and which therefore does not necessarily occur during the initial perception of the external stimuli), could thus be part of the cerebral activity of the brain at resting state (Ansermet and Magistretti, 2010).

These observations have in turn led some scientist to draw a parallel between the function of the self, notably in its

Freudian definition, and the DMN (Carhart-Harris and Friston, 2010; Cieri and Esposito, 2019; Cieri, 2022). This analogy is based on evidence that this network has been linked to spontaneous cognition and self-referential activity that not only involves remembrance of the past and projection into the future but also the processing of interoceptive content. Indeed, the psychoanalytic definition of an example of the psychic self, involves an important physical and interoceptive dimension: according to Freud, the primary formation of the ego, in the phase of “primary narcissism,” would occur with the libidinal cathexis of the subject’s own body, when all the psychic energy unifies and organizes itself taking the ego as object (Freud, 1914). The ego would then become the object of the drive—the sexual drives finding their origin in excitation coming from the bodily organs. Thus, according to Cieri, in so far as in Freudian theory the psychoanalytic significance of the body took on an increasingly central position (the ego, in the second topographic model, is understood primarily as a being a physical entity) its psychic processes appear coherent with the neuroscientific data on brain activity at baseline conditions (Cieri and Esposito, 2019). Furthermore, the ego, according to Freud, is made to contend with different demands, stemming both from the external reality and certain interdicts, but also with the demands of the drives coming from the body (Freud, 1923); something that further supports the parallel with the DMN.

Carhart-Harris and Friston (2010) have proposed integrating this parallelism between the DMN and the binding function of the ego in Freudian theory into the framework of the FEP. On the grounds that the DMN (like the ego according to Freud) processes and integrates the interoceptive and exteroceptive signals, one of the functions of the DMN could be the processing of endogenous excitation and the minimizing of the free energy.

It should however be underlined that drawing this parallel between the Freudian model and contemporary neuroenergetics poses an epistemological difficulty. It is true that in his *Project* (1895a) Freud posited a neuronal model of the psychic mechanism. Yet, his abandoning the publication of that manuscript indicates a renouncing of any ambitions to locate the psychic processes within anatomical structures. In chapter VII of *The Interpretation of Dreams* (Freud, 1900), he formulates this renouncement of a model based on anatomical location, thus,

I shall entirely disregard the fact that the mental apparatus with which we are here concerned is also known to us in the form of an anatomical preparation, and I shall carefully avoid the temptation to determine psychical locality in any anatomical fashion. (Freud, 1900, p. 536)

The proposal to locate the Freudian mechanism of the ego within a cerebral network is therefore to step away from the epistemological project at the origin of psychoanalysis.

Nevertheless, beginning in the 1900’s an important new current of research, “neuropsychanalysis,” proposed the creation of a new discipline. The objective was to put in place an integrative approach. This approach would bring together psychoanalysis and the neurosciences within a new entity, using the methods and the techniques introduced by neuroscientific developments, such as the data from functional imaging. The aim of this new discipline was to form a link between psychic processes described by psychoanalytic theory and the neuronal structures involved in psychopathological disorders. According to Solms and Turnbull, this approach is the only one that gives scientific objectivity to psychoanalysis. This is an objectivity without which psychoanalysis consists purely in a study of subjectivity devoid of any scientific or experimental foundations (Turnbull and Solms, 2007).

The analogy between the function of the binding of the ego’s psychic energy described by Freudian theory and the mechanisms of the DMN as it has been described by Friston et al. fits into this neuropsychanalytic perspective.

In Freudian theory, the ego seeks to convert the free energy into a state of bound energy. As early as his “Project for a Scientific Psychology” (1895a) Freud had posited a main function of the psychic apparatus: the tendency to discharge quantities of excitation stemming both from within and from without the body. The distinction between free energy and bound energy, that Freud will again return to in *The Interpretation of Dreams* (1900a), encompasses in fact both primary and secondary processes that make up two levels of functioning in the psychic apparatus. This division of energy into two categories of differing nature brings to mind the first distinction made by Helmholtz between free energy (that can turn into different kinds of energy), and bound energy (that is manifest solely as calorific). However, in Freud’s writings “free” should be understood as “free to move” (*frei beweglich*) and not as “free to transform” (Laplanche and Pontalis, 1988).

The free flow of the energy corresponds to the primary mechanism of the psychic functioning that Freud had initially called principle of neuronc inertia, in which the energy has a tendency to be discharged in a complete and immediate manner. The secondary process, a psychic mode of functioning that appears in a second phase, involves a binding of the energy. That is to say, the energy is pent-up and retained within certain neurons or neuronal networks, where it accumulates. Dating from 1895, it was to the ego that Freud had attributed the task of binding the psychic energy; of maintaining it constant at a relatively high or at the very least positive level, a level “above zero,” within which a constant level of cathexis is maintained. Hence, the term *Bindung* (binding) indicates at that stage the movement from a state of free energy (where it would flow through the fastest routes leading to discharge, and thus to a zero level), to one of bound energy (that is, maintained at a low but positive level, preventing the immediate flow toward discharge). According to the theory of neurons presented in his “Project for

a Scientific Psychology” (1895a), the binding is made biologically possible through the existence of an network of neurons between which connections could be built: “Now the ego itself is a mass like this of neurones which hold fast to their cathexis—are, that is, in a bound state; and this, surely, can only happen as a result of the effect they have on one another” (Freud, 1895, p. 368).

This binding function of the ego is, according to Carhart-Harris and Friston (2010) consistent with the mechanisms of the DMN. A number of sources argue that one of the main functions of the DMN is the elaboration of internal mental simulations whose purpose is adaptive (Andrews-Hanna, 2012; Buckner, 2013). By integrating this observation into the FEP model, it can be deduced that the system tries to maintain the internal equilibrium of the organism through these mental simulations or representations aimed at adaptation (Carhart-Harris and Friston, 2010). Thus, the DMN, acting as a mediator between the internal and external stimuli is, from the point of view of the FEP, an essential tool for maintaining equilibrium. The energy investment by the system in trying to maintain the lowest levels of entropy (Carhart-Harris, 2018), reduces the risk of encountering the unexpected, and thus minimizes future uncertainty (Friston et al., 2012; Cieri and Esposito, 2019). In Friston’s model of the FEP, each system tries to eliminate the free energy of its subordinate systems, through a process of optimisation of predictions for reducing errors. In this process it is the DMN that is most central (Carhart-Harris and Friston, 2010). The parallel between the Freudian ego and the DMN seems also to be supported by studies that demonstrate that the connectivity between the different nodes in the DMN is absent in newborns, and that it is built and matures during development (Fransson et al., 2007; Fair et al., 2008; Kelly et al., 2009). This could correspond to the Freudian hypothesis of an ontogenetic development of the ego during the primary narcissism stage (Freud, 1914). Furthermore, the hypothesis according to which one of the functions of the DMN (akin to the Freudian ego) could be the integration of interoceptive and exteroceptive signals, can also lead to one to consider that the processes of brain plasticity (by which experiences stemming from our interaction with the environment are inscribed and leave a trace in our neuronal network) contribute to the brain’s expenditure in energy at its baseline condition. These processes could correspond to an “offline” mode of functioning of the brain, to the extent that they involve mechanisms that, even if they are triggered by perception, are temporally out of step with that perception. The mechanisms of plasticity can thus be deployed following a perception even when that perception has ended, since the information continues to be processed by the brain (Ansermet and Magistretti, 2010).

The mechanisms of neuronal plasticity (such as long-term potentiation and depression, or the processes of consolidation and reconsolidation, that unfold several hours to several day after the experience has taken place) call on a great many cellular and molecular processes: the activity of the ionic channels,

the activation of enzymes, protein synthesis, or the regulation of the expression of different genes that contribute to the structural modification of synapses. These processes present a energetic cost to the brain, and that even when it is performing no specific task or activity (Ansermet and Magistretti, 2010). The processes of plasticity can therefore add to the significant expenditure in energy observed during the baseline state, therefore participating in the activity of the DMN.

In this way, if the mechanisms of plasticity contribute to the continued energy activity of the brain in resting state, this view encourages the drawing of a parallel between the energy consumption of the brain in default mode and the functioning of the unconscious. Indeed, in 1925, when Freud proposed a theory of the unconscious as a system of psychic traces stemming from the perceptual experiences, he had specifically pointed to the existence of such an alternating functioning of the psychic apparatus. Periods when conscious perceptual attention was directed on external environmental stimuli would be followed by periods during which this exteroceptive activity would cease. The functioning of the conscious perceptual system tasked with receiving the stimuli coming from the environment would thus be discontinuous, and the traces resulting from the periods of excitation would subsequently be preserved and stored within the unconscious. The unconscious would form, inversely to the perceptual apparatus, a system whose functioning is continuous, where traces would continuously be processed and reassociated with each other following different criteria of association (Freud, 1925). In that regard, the opposition between an unconscious system that functions continuously, and a discontinuous perceptual system focused on the environment (a system that stops during, for instance, daydreaming or sleep), can present a resemblance with the “two views of brain function” developed by Raichle. This is a view that is based on data from functional imaging, where two types of brain activity alternate: one made up of the activation of certain areas during deliberate conscious tasks, the other linked to the activity of the DMN when no perceptual or deliberate activity is taking place (Raichle, 2010).

In a more general way, it remains particularly interesting to consider the hypothesis according to which the perception of somatic states and the integration of interoceptive information combined with information coming from the environment (processed offline by the mechanisms of cerebral plasticity), could constitute one of the components of the energy expenditure of the DMN; as well as to reflect on these data in the light of FEP. This interdisciplinary perspective, aimed at bringing together research on the DMN with concepts stemming from psychoanalysis, shows significant promise with regards to its application to the understanding of certain psychiatric pathologies. Over the past few years, the DMN has received increasing attention and it has come to light that this network undergoes alterations in neurodegenerative illnesses, as well as in certain psychiatric pathologies (Cieri et al., 2020). More

specifically, the data coming from neuroimaging attest to the existence of a hyperactivity of the DMN in schizophrenia. These data are particularly interesting when seen through the prism of the energetics concept of the psychic processes developed by Freud (notably his description of the libidinal processes at work in psychosis), and when we consider them within the FEP.

The hyperactivity of the default mode network in schizophrenia

Since publication in 2001 of Raichle's first article on the baseline state *A Default Mode of Brain Function* (Raichle et al., 2001), close on 3,000 studies have been done on this subject in laboratories around the world. Among these studies, work focusing on the relationship between the functioning of the baseline state and psychiatric or neurological illnesses represent by far the largest proportion (Raichle, 2015). There exists an anomalous functional connectivity of these networks in several mental disorders, such as anxiety, depression, bipolar disorder, OCD, autism, and Alzheimer's disease (Andrews-Hanna et al., 2014). Buckner (2013) has underlined the importance of studying the DMN for psychosis, stating that "Psychosis may be a network disturbance that manifests as disordered thought, partly because it disrupts the fragile balance between the default network and competing brain systems." As regards the study of alterations of the baseline network state in schizophrenia, many studies have highlighted the existence of a hyperactivity of the DMN in patients with schizophrenia when compared with healthy control groups. There are, however, studies in the literature that reach the opposite conclusion (Zhao et al., 2018; Lee et al., 2019). An absence of concurrence comes out of these studies. The results remain difficult to interpret, notably because of a significant variability in symptoms between individual (Lafargue, 2010).

Based on a study of MRIs focused on the baseline state in schizophrenia patients suffering from auditory hallucinations, Mallikarjun and his team have suggested that these divergences in results could be explained by differences observed between patients who have recently had their first psychotic episode and chronic patients who have been treated with neuroleptics for a number of years. In order to reduce these biases, recent studies have sought to distinguish between cohorts of patients who have recently decompensated and have thus not benefited from treatment and chronic patients who have received treatment for many years. Proposals have also been made to differentiate patients according to the nature of their symptoms. Some studies have also included, in addition to patient groups and control groups, groups of first-degree relatives of schizophrenia patients, or subjects who present an increased risk factor for schizophrenia but who have no apparent symptoms. A review of the data furnished by MRI studies of the DMN in patients with schizophrenia has highlighted that the majority of studies

observe an increased functional connectivity in the brain activity in these patients—including in chronic patients and patients receiving treatment (Hu et al., 2017). However, three studies found opposite results, observing a reduction in functional connectivity in resting state (Bluhm et al., 2007; Zhao et al., 2018; Lee et al., 2019). All other studies mentioned by the review evidenced the existence of a significant hyperactivity of the DMN in schizophrenia patients (Liu et al., 2012), be it at the onset of illness or in chronic cases (Zhou et al., 2007). This hyperactivity has also been observed in individuals who present high risk factors for psychosis, as well as in the siblings of schizophrenia patients who do not themselves present signs of psychosis (Shim et al., 2010). The schizophrenia patients present an increased brain activity and an increased connectivity at resting state within the dorsolateral prefrontal cortex (Guo et al., 2017), the frontal gyrus, the insular cortex and the dorsolateral cortex, the medial and superior temporal gyrus, and the posterior cingulate cortex (Galindo et al., 2018). Their healthy siblings or close relatives present increased activity within the cingulate cortex and the anterior prefrontal cortex (Guo et al., 2017; Galindo et al., 2018).

Additionally, the schizophrenia patients also present anomalies in the usual deactivation of the DMN during the performance of tasks focused on the environment or requiring high levels of attention, when compared with the control groups (Hu et al., 2017). The activation of the attention system networks when a task where attention is focused on external stimuli is performed, usually involves a deactivation of the DMN (these two networks are termed negatively correlated). This negative correlation can be explained by a competition between the attention resources that are either focused on the exterior or the interior, depending on the need for responses to stimuli coming from the environment (Fransson, 2005). These data can, as we have seen, present some similarities with the Freudian description of the conscious perceptual system's discontinuous mode of functioning; attention being, according to Freud, only punctually outwardly focused, as though consciousness "stretches out feelers, through the medium of the system Pcpt.-Cs., toward the external world and hastily withdraws them as soon as they have sampled the excitations" (Freud, 1925, p. 231). While the rest of the time, psychic activity is dominated by internal activity during which representations stemming from mnemonic traces left by the external excitations can be reassociated with each other.

The model of negative correlation between the DMN and the Task Positive Network (TPN) brought to light by recent research in neuroscience appears particularly impaired in schizophrenia patients. This is noticeable in patients suffering from paranoid schizophrenia (Zhou et al., 2007), and in those with dominant positive symptoms (Garrity et al., 2007), as well as in some first-degree relative of schizophrenia patients (Whitfield-Gabrieli et al., 2009). The functional hyperactivity of the DMN in schizophrenia patients would thus be present

both at resting state as well as during the performance of tasks; and, in general, the patients whose pathological symptoms are the most significant are those who present the lowest level of deactivation of this network during tasks. According to the authors, this defect in the deactivation of the DMN during positive tasks could explain the attention and cognitive disorders observed in some schizophrenia patients (Whitfield-Gabrieli et al., 2009). Since the DMN is usually associated with the presence of thoughts independent of external stimuli—for example introspection and self-referential thoughts—it seems coherent that results, obtained during tasks that require a high level of attention focused on the environment, should be less good when the internal activity is high. This would be because the high level of internal activity could contribute to deficiencies in terms of vigilance and attention owing to competition with the self-referential cognitive activity (Whitfield-Gabrieli and Ford, 2012), for example when a significant delusional activity is present.

The significant number of studies having observed functional hyperactivity and hyperconnectivity within the different areas of the DMN (be it in resting state or during the performance of tasks) in patients with schizophrenia, has led Hu and Zong to state that this discovery represents some of the most important neurobiological data to emerge from recent research on schizophrenia:

On the basis of recent evidence, functional hyperconnectivity within the DMN is perhaps the most common finding in comparisons of schizophrenia patients with healthy controls. It has been extensively found in different schizophrenia populations, including chronic paranoid schizophrenia, first episode schizophrenia, individuals at ultra-high risk for psychosis, and the first-degree relatives of patients with schizophrenia. (Hu et al., 2017)

According to Liu and his team, the hyperactivity of the DMN may more specifically constitute an endophenotype of schizophrenia, but other members of the population could present with a hyperactivity of the DMN without presenting psychotic symptoms (Liu et al., 2012).

Furthermore, if in the majority of studies this hyperactivity is observed, these conclusions should be put into perspective by the extremely variable nature of the results. To explain these differences, some scientists have tried to establish a correlation between the nature of the symptomatology the schizophrenia patient presents with and the anomalies of activity of the DMN. The presence of auditory hallucinations (already linked to activity of the areas involved in auditory processing, language and memory, as well as the salience network) has been linked to hyperactivity of several areas of the DMN. A study by Mallikarjun has shown that the superior temporal cortex, the insula, the precuneus, the posterior cingulate cortex, and the

parahippocampal region, are particularly active areas in subjects who have presented auditory hallucinations during the fMRI sequence (Mallikarjun et al., 2018). The patients who hallucinate show an increased functional connectivity at rest between parts of the salient network and the DMN, but a reduced functional connectivity between the claustrum and the insula, compared to the healthy cohort. Scientist have expressed the hypothesis that the hyperactivity observed within the DMN, and the salience network, could be linked to an alteration in the evaluation of the salience of the internal activity, an alteration that would then contribute to the origin of the hallucination.

Although the function of the claustrum remains poorly understood, some studies have brought to light that this structure may be involved in consciousness of perceptual experiences. Anomalies of the claustrum could compromise the synchronization between the associative cortex of the prefrontal lobe and the parietal lobe (Koubeissi et al., 2014; Yin et al., 2016; Fodouliau et al., 2020).

Another similar theory of the function of the claustrum posits that this brain region is involved in the timing and coordination of the cortical activity that results ultimately in perceptual experience. Disruption of the binding or synchronization of sensory, cognitive, and motor information could contribute to delusions and hallucinations in schizophrenia patients. There is a possible further link between the severity of delusions in the disease and structural abnormalities in the claustrum (Casella and Sawa, 2014).

It has also been proposed that the diminution in functional connectivity observed between the claustrum and the insula could lead to a compensatory overactivation within some areas of the auditory network, including in some areas of the DMN and areas associated with the processing of language and memory (Mallikarjun et al., 2018).

These data regarding an increased activity of the DMN in patients who present with auditory hallucinations support the results from studies that have demonstrated that the hyperactivity in the DMN could be attributable to what is termed the “positive” symptomatology of schizophrenia (Garrity et al., 2007). It has also been demonstrated that, inversely, a correlation exists between negative symptoms and the activity of the right anterior prefrontal cortex at resting state (Reichenbach et al., 2012). The increased hyperactivity is particularly important in schizophrenia patients who present with delusional ideas with a paranoid thematic, when compared with other patients (Zhou et al., 2007) and, as we have seen, in patients suffering from auditory hallucinations (Garrity et al., 2007).

The positive symptoms of schizophrenia according to the Freudian perspective

Within the scope of a dialogue between psychoanalysis and neuroscience centered on an energetics concept of mental

activity, and in the light of a psychoanalytic understanding of psychosis, it is particularly interesting to deliberate the data coming from neuroimaging, in particular data that bring to light the existence of an increased hyperactivity of the DMN in patients with schizophrenia who present positive symptoms (characterized by delusional ideas and auditory hallucinations). To do this, we are first going to evoke the Freudian hypotheses regarding the positive symptoms of schizophrenia, particularly from the perspective of delusions in psychosis understood as an “attempt at recovery.”

Freud, as we have seen, had proposed early on in his work a concept of mental disorders based on energetics. Once he had abandoned references to physiology and the anatomy of the nervous system, it is the concept of libido that becomes the focal point of the economics of his metapsychology. Starting with his study of the Schreber case, Freud narrowed down his understanding of the libidinal processes at work in psychosis. In the context of his disagreement with Jung, he remodeled his definition of libido. Freud's research led, in *On Narcissism: An introduction* (1914c), to the distinction between narcissistic libido which cathects the ego and an object libido turned toward the outer world. The introduction of narcissism would then enable Freud to disentangle the two phases of the psychotic process that consisted in a twofold libidinal movement. The first phase of psychosis consisted in a massive withdrawal of libidinal cathexis, of all the psychic energy, from external objects; the libido flowing back, exclusively on the ego and the subject's own body. This first phase was characterized by the fact that the patient (notably in those with schizophrenia) “...withdraws his interest from the external world completely...” (Freud, 1911, p. 75). The negative symptomatology of schizophrenia, such as being withdrawn, blunted affect, apragmatism, and also deteriorations in representations of the body (cenesthetic hallucinations, hypochondria, dysmorphophobia) are characteristic of this initial phase.

With Schreber this initial libidinal phase was expressed notably by a world's end fantasy, in which the universe is perceived by the subject to be totally devastated and depopulated, with men reduced to being only botched shadows:

The patient has withdrawn from the people in his environment and from the external world generally the libidinal cathexis which he has hitherto directed on to them. Thus everything has become indifferent and irrelevant to him... The end of the world is the projection of his internal catastrophe; his subjective world has come to an end since his withdrawal of his love from it. (Freud, 1911, p.70)

Freud postulated that this first phase is common to all forms of psychosis. However, in order to fight against this movement of libidinal withdrawal, some patients put in place a second process that would consist in an attempt to reinvest external

objects with the libido. This second phase found its expression in productive symptomatology such as auditory hallucinations and delusional ideas:

When he does so replace them, the process seems to be a secondary one and to be part of an attempt at recovery, designed to lead the libido back to objects. (Freud, 1914, p. 74)

With Schreber, the paranoid delusion thus formed an attempt to “reconstruct” the world, for after having withdrawn his libidinal cathexis from the external objects of the world in the withdrawal phase, he sought to reinvest objects through his delusion. It was in effect a delusional external reality he constructed but, from a strictly libidinal perspective, this motion did constitute an attempt to reinvest an objective world: “We take to be the pathological product is in reality an attempt at recovery, a process of reconstruction.” (Freud, 1911, p. 71).

The logic of delusion as attempt at recovery should thus be understood primarily as relating to a libidinal and energetic dynamic, and in the diachronic perspective of a succession of two phases within the psychotic process. Freud's formulation therefore puts the emphasis on the essentially temporal character of psychosis. The delusional construct unfolds progressively, as Schreber's autobiography demonstrates, and can sometimes spread over an entire lifetime. In this, Freud is breaking with theories that see delusion as a deficiency; in a gesture that is seen as radically subversive by classical psychiatry. Thus, he proposes assimilating the delusion production to an “autoplastic” creative act (106), similar to an *active phase* involving a high cost in energy since through this delusion process “the ego creates, autocratically, a new external and internal world;” (Freud, 1924a, p. 151).

Furthermore, Freud argued that the process of delusional reconstruction implied not only the creation of a new reality but also required the psychotic to rebuild their memories, that is to say “...the store of memories of earlier perceptions...” (Freud, 1924a, p. 150) that constituted a copy of the past world. Thus, the reconstruction not only necessitated the generation of new current perceptions, perceptions through which “Normally, the external world governs the ego...” (Freud, 1924a, p. 150) but must also, in order to be operant, modify past perception, as though they were witnesses to a previous world, which must be erased:

In psychosis, the transforming of reality is carried out upon the psychical precipitates of former relations to it – that is, upon the memory-traces, ideas and judgements which have been previously derived from reality and by which reality was represented in the mind. (Freud, 1924b, p. 185).

The scope of the delusion reaches even into memory itself, remodeling the mnemonic traces despite them being the traces of

past perceptions. Now, if according to Freud the memories of these past experiences form the main basis of our relationship to the real world, this relationship is nevertheless not closed and is "...continually being enriched and altered by fresh perceptions." (Freud, 1924b, p. 185). The unconscious work of this psychotic rebuilding therefore also has as its objective, the creation of perceptions that can correspond to a new reality, aim that would be "...most radically effected by means of hallucination." (Freud, 1924b, p. 186).

Freud, therefore, sees auditory hallucination as being an integral part of reconstruction of the delusional world in a dialectic between memory and perception: if the delusion rebuilds the past traces of the relationship to reality through mnesic illusions of false memories, it also reconstructs the current relationship through the creation of new perception (Freud, 1924b). Auditory hallucinations should thus be understood as *mad traces* (Assoun, 2005). Indeed, for Freud, hallucination does not belong to a perception that is false or without object but is to be seen more specifically as a confusion between a memory and a current perception (Freud, 1923), where the memory finds itself invested with the strength that usually characteristic of the cathexis of external perceptions.

The mechanism through which the delusional reconstruction operates, from the Freudian perspective, generally involves a repression of all the mnesic network. The creation of a new reality does not come about *ex nihilo* but happens on the basis of mnesic traces that have really been perceived. Traces that are then reassociated and reorganized between themselves in new ways.

Hyperactivity of the DMN in schizophrenia and the reassociation of traces during delusions

The Freudian psychoanalytic perspective according to which the positive symptomatology of psychosis (delusional ideas and auditory hallucinations) is the result of a process of trace reassociations, is interesting to put into perspective alongside the neuroimaging data that evidence a hyperactivity of the DMN in schizophrenia patients, in particular those for whom positive symptoms are predominant. Nowadays, this interaction between reassociation of traces and DMN might be accounted for within the framework of the free energy principle (FEP) (Rabeyron and Massicotte, 2020), and various forms of failure of the process of linking, regulation and transformation of free energy within the psychic apparatus could be considered as the origin of different psychopathological manifestations (Rabeyron, 2021).

As we have seen, the activation of the DMN is generally associated with spontaneous cognitive activity and self-referential processes; in particular the remembrance of autobiographical memories and projections into the future

(Vincent et al., 2006; Andrews-Hanna et al., 2010; Doucet et al., 2011). The activation of the DMN in the absence of a task focused on the environment (Andrews-Hanna et al., 2010) appears to be particularly dominated by the recollection of autobiographical memories (Doucet et al., 2011). These processes, as we have seen, involve a neuronal network that belongs to the DMN, one that associates certain regions of the parietal cortex and regions of the hippocampus (Gusnard and Raichle, 2001; Vincent et al., 2006). We have seen that the medial parietal cortex, as well as the medial prefrontal cortex, are themselves the seat of hyperactivity at resting state in patients with schizophrenia. Neither do these areas deactivate, as much as in the healthy control group, during the performance of a specific task.

Thus, one can make the hypothesis that the hyperactivity of the DMN in psychosis constitutes the physiological expression of the massive reassociation of traces taking place as part of the psychotic process. This is a process that, notably, involves autobiographical memories. The participation of brain areas which are part of the DMN in autobiographical memory and in the planification of future event, two aspects which are essential to the generation self-representation (what Damasio terms autobiographical self), has encouraged scientist to make the link between changes to self-representation in schizophrenia and the presence of an excessive cerebral activity at resting state. As Lafargue underlines:

The function of the DMN is probably to ensure the coherence of thoughts throughout life, be it memories or imagining the future, by facilitating the elaboration of mental models linked to personal events. This function of cognitive oversight appears to be altered in schizophrenia. Anomalies in the functioning of the DMN have been brought to light (Lafargue, 2010, p. 47)[our translation]

If, as we have seen, the productive symptomatology may stem from a massive reassociation of traces, hyperactivity of the DMN would thus appear to be the physiological outcome of this process. The proliferation of representations at the onset of the process of reconstruction of the world performed by the delusions [prodrome that the alienists had described as an "incubation period" of the delusional ideas (Sérieux and Capgras, 1982)] might manifest itself through an increased activity of the spontaneous cognitive processes involving, and therefore an increased expenditure of energy in the areas associated with the DMN. The hyperactivity observed in the DMN in schizophrenia particularly affects the posterior regions (such as the posterior cingulate cortex and the inferior parietal cortex) and the anterior regions (medial prefrontal cortex, hippocampus and lateral temporal cortex). These are areas that are almost identical to those activated during experiments that require the subject to recall certain autobiographical memories (Svoboda et al., 2006), and encompass specifically

the parietal-hippocampal network that Vincent and his team describe as particularly involved in episodic memory and the recollection of memories. The function of the delusion according to Freud focuses on "...the store of memories of earlier perceptions..." (1924b, p. 150). In order to rebuild reality, the delusional construct must, even before creating new perceptions, modify the *reflection* of the external world that are our representations stemming from mnemonic traces left by past perceptions. If the *remodeling of reality* focuses selectively on "...the memory-traces, ideas and judgements which have been previously derived from reality" (Freud, 1924b, p. 185), we may conceive that, from a neurobiological perspective, this reconstruction makes significant demands on the regions of the brain involved in autobiographical memory.

Furthermore, the areas situated in the medial line of the DMN are also activated when the subjects anticipate or imagine future events (Addis et al., 2007). This aspect plays a significant role in delusional constructs. Specifically, anticipation of events to come appear to be one of the major components of most delusional thematics. Delusions of persecution are characterized, in their most severe instances, by a massive anxiety on the subject's part regarding their future, even going as far as believing that an attempt on their life is imminent. With erotomania, it is generally the prospect of marriage or of imminent reunions with the loved one that dominates the representations linked to the future. This is the case with the patient Freud describes in his article *The Neuro-Psychoses of Defence* (1894a), who is convinced that her beloved, who had abandoned her, will return to her at a specific date. Similarly in the final evolution of Judge Schreber's delusion, where peace seems to come from the fact that he defers the realization of his transformation into a woman to a distant future (Schreber, 2000). Representations linked to projections into the future are thus particularly solicited by the delusional activity, and that even within the different thematics of delusion.

Finally, studies on the correlation between the different forms of spontaneous cognition and the activation of the DMN have brought to light that the temporoparietal junction, the cingulate cortex and the precuneus (areas that are particularly active at resting state in schizophrenia patients) (Saxe and Kanwisher, 2003; Spreng et al., 2009) intervene not only in autobiographical memory but are also activated when the participants try to imagine a situation through the eyes of someone else, "putting themselves in another person's shoes." What cognitive psychology terms theory of mind—that is to say, the ability to project oneself into, and imagine what is happening in, the mind of another person. Now this cognitive process is particularly involved in the construction of delusional ideas—notably in paranoid ideas, where others are imputed with malevolent ideas.

The hypothesis of a strong involvement of spontaneous cognition—more specifically autobiographical memory, the representation of future events, and theory of mind—in the activation of the DMN, brought to light by several studies

of imaging, looks to be particularly promising for explain the existence of cerebral hyperactivity at resting state in schizophrenia patients. As Bastin underlines:

In fact, all the tasks that activate the DMN require that the individual project themselves into a situation other than the reality that surrounds them. They abstract themselves from the environment to build a mental model representing alternative scenarios. Indeed, rather than process external stimuli, cognition is turned on the self, generating personal memories, imagining possible future events, and reflecting on personal emotions and motivation, as well as those of others. It is then, a question of generating mental content oneself. (Bastin, 2018) [our translation]

It is precisely this capacity to generate and construct new scenarios, separate from external reality, that is significantly mobilized in the process of delusional reconstruction such as Freud envisaged it in the form of the construction of a new reality. The "autoplastic" (Freud, 1924b) creations of psychosis—where the subject brings together reassociations of traces that form a new assembly of representations, discontinuous from the initial perceptual experiences—can thus be characterized as belonging to that type of internal cognition that is self-referential, focused on the self, and generating personal memories that are modified, reassociated, then put into perspective with representations of some future events. The hyperactivity of certain regions of the DMN observed in schizophrenia—particularly within the prefrontal cortex, the precuneus and the parietal areas—, in so far as that it affects the areas involved in episodic memory, the recollection of autobiographical memories, the representations of other people's somatic states, and the ability to generate mental simulations as well as projection into the future, could therefore be correlated with an important activity of these different cognitive functions during the process of constructing a new reality by way of delusions. This hypothesis can be underpinned by the fact that scientists have frequently noted a correlation between this hyperactivity and the presence of a positive symptomatology in the schizophrenia patients (Whitfield-Gabrieli et al., 2009).

Discussion

We will now look at integrating this position with the FEP, which assimilates the functioning of the brain to the stationary state of a non-equilibrium system (Jarzynski, 1997) while also drawing on a Bayesian approach (Cieri et al., 2021). According to Friston (2009), the brain strives to minimize variations in free energy by maximizing evidence drawn from perceptual experiences and by reducing the element of surprise through an active role on the outside world (Ortega and Braun, 2010).

Freud speaks of an energy quantity that can be either free or bound. This is quite distinct from the formal notion

of free energy of thermodynamics, statistical physics and active inference (Parr et al., 2022). However, we observe that psychoanalysis and the neurosciences agree when it comes to the observation that some representations—particularly those stemming from our experiences with the environment, which are stored in our autobiographical memory—can work as a medium for the binding function. These representations take on, in combination with traces from somatic states, a homeostatic role in the psychic balance and are generators of a form of mental equilibrium (Freud, 1900; Ansermet and Magistretti, 2004, 2010; Tran The et al., 2021).

In a FEP perspective, we can assume that the brain tries to maintain an internal equilibrium based on representations constructed from traces that stem from our past experiences. These traces are associated with our internal physiological and somatic states. As Solms has suggested, this minimizing of the free energy can be put in parallel with a homeostatic understanding of the regulation of physiological states (Solms, 2013, 2019; Solms and Friston, 2018). This is a point of view that is also held by Damasio (2000). From this standpoint the DMN, acting as mediator between the internal and external stimuli, is an essential tool for the maintaining of an equilibrium aimed at minimizing free energy (Cieri and Esposito, 2019). Specifically, the superior cortical areas try to organize the activity of the lower levels by decreasing their free energy (Carhart-Harris and Friston, 2010).

The French psychoanalyst Jacques Lacan took a particular interest in the idea that the traces that stem from the inscription of the Oedipal experience could have similar effects of balancing the psyche (Lacan, 1946). Likewise, certain traces (conscious or not) and autobiographical memories that are essential to the generation of what Damasio terms our consciousness of an autobiographical self (which are at the basis of our sense of self identity and which are stable through time), can be part of this maintenance of equilibrium in energetic terms (Damasio, 2000). This energy equilibrium might be linked to Noether's physics theorem according to which a system that has time translation symmetry is characterized by energy conservation.

It is probable that this process of equilibrium fails in the case of schizophrenia. Following the hypothesis of disconnection, this pathology is characterized by an anomalous interaction between the specialized cerebral regions. This results in a defective integration of the activity between the different networks, and thus a breakdown in cognition. With this pathology there is also a failure in the process of inference, owing to a defect in the modulation of somatic states (Friston and Frith, 1995; Friston et al., 2016). This build-up of disorder can be associated with what Freud had characterized as the primary phase of the psychopathological process at work in psychosis, a process dominated clinically by disorders in perception of the body and in interoception (Freud, 1911), which constitute the precursory stage to schizophrenia (Freud, 1915–1916/1963). This coincides with the presence in this pathology of an

increased sub-cortical hyperconnectivity that translates into a heightened entropy within the brain as a whole (Salman et al., 2019). The result of psychosis would be a failure of the process of minimization of free energy and, statistically speaking, any exteroceptive or interoceptive sensation could then become surprising and unpredictable. Thus, according to Carhart-Harris and Friston (2010), in schizophrenia the functions coming down from the brain fail to predict errors. This then translates into an increase of the free energy which cannot be controlled. If we put this hypothesis alongside the role of binding free energy usually performed by certain traces and autobiographical representations, we can argue that this binding process appears to be jeopardized in schizophrenia.

This hypothesis can also be put alongside the research that has brought to light lacunae amongst the neurons involved in long-term memory and memory consolidation, in schizophrenia patients; in particular, lacunae within certain populations of neurons with precocious neurogenesis that express parvalbumin within the hippocampus (Donato et al., 2015; Genzel et al., 2015; Carvalho, 2017). These lacunae could lead to a lack of consolidation and stability in some traces and in autobiographical memories, which are essential to the generation of what Damasio terms our awareness of an autobiographical self that is at the root of our sense of self identity (Damasio, 2000). In the absence of stability over time, these traces could perhaps be more subject to lability and reassociations in psychotic patients.

Indeed, according to Bastin, the cerebral circuits of the DMN habitually “allow individuals to experience identity and consciousness” (Bastin, 2018) [our translation]. The unstable aspect of identity and of autobiographical consciousness of self in psychosis (notably because of the absence of a fixed stability of the traces that are the basis of this identity) could thus create massive functional reorganization of the DMN's activity. This neuroscientific hypothesis of a link between the hyperactivity of the DMN and a massive spontaneous cognitive activity is particularly rich in possibilities, if it is looked at through the prism of the Freudian definition of delusion as an attempt at recovery. Where the traces that make up our sense of autobiographical self usually appear to be stable and well-organized, thus guaranteeing low entropy, their lack of consolidation in schizophrenia may have as consequence a state of disorder, away from equilibrium (or from homeostasis) and an increase in free energy fluctuations. The disruption of these mnemonic traces or representations that are at the root of identity and a sense of self, might then be assimilated to a dissipative process, i.e., to an internal production of entropy (Prigogine, 1967; Prigogine and Stengers, 1984). This could correspond to the presence of a globally and locally more entropic IRMf signal among schizophrenia patients. This increase in cerebral entropy constitutes an example of the disorganized peak activity that underlies schizophrenia (Sokunbi et al., 2014). According to Carhart-Harris and Friston (2010) acute psychosis corresponds

to a mode of functioning described by Freud as the primary process, where the free energy is unbound. The DMN, which normally functions to minimize free energy through a hierarchy that goes from the lower to the higher levels, fails in its task. The FEP, which understands cognition as based on Bayesian hierarchical inferences and time evolution of the free energy near a stationary state, requires that the superior cortical areas work to minimize the free energy of the superior levels (Mumford, 1992; Rao and Ballard, 1999; Kiebel et al., 2008; Friston, 2010). At the onset of psychosis, the DMN appears unable to control, in a descending manner, the excess of free energy in the lower levels. More specifically, areas of the DMN such as the ACC are dysfunctional in schizophrenia, and these anomalies have also been linked to disfunctions in the anterior insula (Carter et al., 2001). Furthermore, the data from neurobiological literature on anomalies of the insular cortex in schizophrenia, have demonstrated that the disfunction of the insula can constitute one of the underlying biological causes of the disorders in corporal perception present in schizophrenia (Tran The et al., 2021).

According to the FEP, the brain uses internal hierarchical models to predict entering sensory perceptions: the brain constructs its own descending predictions (from cortex to perception) on the basis of sensory samples of the world (Friston, 2009). A failure of the integrating function of the DMN, between the exteroceptive and interoceptive stimuli, can create an inability to construct adequate descending predictions. Thus, the aim of the process is to optimize the explanation of the causes of sensory inputs, in order to establish new predictions that inform action and behavior (Friston, 2003). In schizophrenia, this process fails, and surprise (and thus the fluctuations of the free energy) will increase.

Following the Freudian hypothesis of delusions as attempts at recovery, the process of systemising the delusions can constitute an attempt by the organism to try and minimize the free energy and the disorder present in schizophrenia. This would be an instance of the tendency of all organized systems, and notably living organisms, to seek to lower the level of surprise as it interacts with its environment (Friston et al., 2015). The representations fixed by the systematization of delusions could represent an attempt to minimize the free energy, of binding it through the consolidation of the delusional representations in a process similar to that which Hopkins (2016) describes. This is a process based on the normal function of dreams, which perform a reduction of complexity (a process of binding) through the consolidation and reconsolidation of memory. Since these mechanism for reducing complexity appear to play an important role in mental disorders (Hopkins, 2016), it can be assumed that delusions are an attempt at converting free energy into bound energy through the production and the consolidation of new mnemonic traces or representations. These traces or representations stem from the processes of reassociation described by Freud. According to

the Bayesian approach, the systematizing of delusions generates new predictions aimed at reducing surprise and re-establishing a form of descending causality. This process echoes Freud's notions of free vs. bound energy. It is consistent also with the notion of an allostatic equilibrium of psyche.

This bringing together of the Freudian idea of delusion as attempt at recovery and Friston's FEP has the advantage not to reduce the neurobiological processes at work in psychosis to a deficit but rather as a functional reorganization. Indeed, even studies that have brought to light the existence of some deficits in anatomical connections in regions of the cerebral cortex among patients with schizophrenia argue that the functional hyperconnectivity between these regions could represent a neuronal response to that an anatomical deficit (Skudlarski et al., 2010). The dialogue between Freudian theory and neuroenergetics, therefore, reintroduces a dynamic perspective into the neurobiological study of the physiopathology of schizophrenia. This goes beyond an exclusively localizationist perspective reduced to a strictly anatomical view.

To envisage that the formation of delusional traces and representations, and the systematization of delusion, might be attempts to restore a homeostatic equilibrium by minimizing the free energy, opens up interesting new therapeutic horizons for the psychotherapeutic treatment of schizophrenia. The observation that there exists a defect in the consolidation of long-term memory in this pathology (Donato et al., 2015; Genzel et al., 2015; Carvalho, 2017), might suggest that the psychotherapeutic work with psychotic patients could have for its objective an alternative process to the delusion's function of binding free energy. The aim would be that the psychic treatment based on talking would enable a reinscription or a reconsolidation of those traces. This would take place through a process of remembering of traces linked to the patient's experiences and their personal history. The hope would be that this would stabilize elements of the autobiographical memory, which could contribute to a greater "sense of self identity," as suggested by Damasio (2000). One could imagine that regular and repeated psychotherapy sessions could stimulate and stabilize this sense of identity and consciousness of self. The consolidation of autobiographical traces would operate a kind of binding of the free energy (or minimizing of the errors in prediction according to FEP), and a regulation of the alteration in the perception of somatic states present in psychosis.

Conclusion

To conclude, we can say that psychoanalysis and neuroscience can find a point for discussion around the inference models of mental activity expressed in terms of a free energy principle, and that this dialogue may be particularly fruitful when it comes to the understanding of some psychiatric pathologies. We have observed that the Freudian

understanding of psychosis is framed by a theory of libido and, thus, of the energetics of the psychic processes. In particular, the hypothesis of delusion as an attempt to heal involves understanding psychosis as a libidinal (therefore energetic) and diachronic process in two phases. Furthermore, Freud conceived the genesis of delusional ideas in terms of a process of “reconstruction” of reality. This is to be understood as a process that operates through the formation of new traces by means of a massive reassociation of traces stemming from past perceptual experiences. This process then introduces a significant discontinuity between the initial experiences and the new traces (delusional ideas and hallucinations) produced by the individual.

Furthermore, much of the data highlight a hyperactivity of the DMN in schizophrenia, particularly for patients in whom the positive symptoms dominate the clinical picture. In so far as studies have shown that the activation of the DMN is more specifically linked to processes of spontaneous cognition and self-referential thoughts (such as recalling autobiographical memories or as projections into the future) this hyperactivity of the DMN can be put in parallel with the Freudian concept of positive symptoms in psychosis.

In the perspective of a dialogue between psychoanalysis and neuroscience, we have thus proposed to consider the data showing a hyperactivity in the DMN in resting state in patients with schizophrenia as being correlative to a massive process of reassociation of mnemonic traces. This is a process that would result in the formation by the psychotic individual of new assemblies of traces and representations, from which delusional ideas or auditory hallucinations would originate. This production of new assemblies of traces would involve those areas, termed “spontaneous cognition,” that are associated with the activity of the DMN, in particular autobiographical memory but also anticipation of future events. Indeed, the process of delusional constructions bears on the one hand on the traces resulting from past perceptual experiences, which it tries to remodel; on the other hand, it results in the generation of new representations of the individual’s future. In addition, this production of new representations introduces a reconfiguration not only of the representation of the self but also representations of others and their intentions. This also joins the imaging data relating to the contribution of mental activity linked to theory of mind, on the activation of the DMN.

Following the FEP model, it is possible to envisage a function of the DMN that aims to minimize the free energy, drawing a parallel with the binding function of the ego in Freudian theory. The creation and consolidation of new representations during the process of delusional systemisation could, in that case, be thought of as an attempt at minimizing the excess of free energy present in schizophrenia. This hypothesis is in keeping with the Freudian concept of delusion in psychosis as an attempt at recovery.

If an interdisciplinary dialogue has been shown to be rich in material for better understating the symptoms and psychopathology of certain psychiatric illnesses, the data regarding the hyperactivity of the DMN still need to be better understood. Indeed, the functions of the DMN are still subject to debate in the scientific community, and some studies done with schizophrenia patients have presented sometimes contradictory results. Nevertheless, integrating the ideas stemming from psychoanalysis within the neuroscience research on schizophrenia may help understand psychosis as a diachronic process. This, in turn, could encourage the setting up of longitudinal studies to better understand the outcomes and evolution of the clinical picture for long-term patients. The Freudian understanding of psychosis in particular emphasizes the temporal and diachronic logic of this pathology, in the evolution of its libidinal—and thus energetic—dynamic. Integrating this view with research in neuroscience can thus lead to encouraging the development of longitudinal studies in imaging, in order to bring to light the evolution of the activation and the functional connectivity of the DMN at different stages of illness. If, in order to avoid biases (linked for example to the administering neuroleptic treatments to patients over several years), different studies apply protocols involving cohorts of untreated patients who have recently decompensated, it could be still more informative to be able to repeat the study of imaging on these patients at different stages of their treatment. This, in order to determine if their clinical evolution correlates with variations in the level of activity and connectivity of the brain’s DMN. Performing such longitudinal studies could fit in with the essentially diachronic character of the psychotic process as evidenced by Freud in his concept of the twofold process at work in psychosis.

In their review of the neuroscientific literature relating to the activity of the DMN in schizophrenia, Hu and his team specifically encouraged the development of such longitudinal studies:

...because the divergence of findings among studies could represent differences in the activity and connectivity of the DMN that might be involved in different stages of this disease, it could be important to explore the DMN throughout the course of schizophrenia... (Hu et al., 2017)

The study of the evolution in levels of baseline brain activity in longitudinal neuroimaging protocols on psychotic patients over several years, from the onset of psychosis up to an eventual clinical stabilization, could represent a new research route. This is a route that would make it possible to take into consideration the psychotic process in its temporal and diachronic dimension, eventually perhaps bringing to light correlations between the levels of activity in the DMN and the patients’ clinical pictures.

Author contributions

JT is the main contributor of this paper as part of her postdoctoral work. FA and PM as supervisors, contributed to the conception and development of the research, and they revised critically the manuscript for intellectual content. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45, 1363–1377. doi: 10.1016/j.neuropsychologia.2006.10.016
- Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *Neuroscientist* 18, 251–270. doi: 10.1177/1073858411403316
- Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335. doi: 10.1152/jn.00830.2009
- Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* 1316, 29–52. doi: 10.1111/nyas.12360
- Ansermet, F., and Magistretti, P. (2004). *À chacun son cerveau: plasticité neuronale et inconscient*. Paris: Odile Jacob.
- Ansermet, F., and Magistretti, P. (2010). *Les Enigmes du plaisir*. Paris: Odile Jacob.
- Assoun, P. L. (2005). La trace folle. Pour une métapsychologie de la trace. *Che vuoi ?* 1, 83–94. doi: 10.3917/chev.023.0083
- Bastin, C. (2018). Le réseau cérébral par défaut: un repos qui n'en est pas un. *Revue de neuropsychologie neurosciences cognitives et cliniques* 10, 232–238. doi: 10.3917/rne.103.0232
- Bloom, R. L., Miller, J., Lanus, R. A., Osuch, E. A., Boksman, K., Neufeld, R. W., et al. (2007). Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: anomalies in the default network. *Schizophr. Bull.* 33, 1004–1012. doi: 10.1093/schbul/sbm052
- Buckner, R. L. (2013). The brain's default network: origins and implications for the study of psychosis. *Dialogues Clin. Neurosci.* 15, 351–358. doi: 10.1387/DCNS.2013.15.3/rbuckner
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Carhart-Harris, R. L., and Friston, K. J. (2019). REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacol. Rev.* 71, 316–344. doi: 10.1124/pr.118.017160
- Carhart-Harris, R. L. (2018). The entropic brain - revisited. *Neuropharmacology* 142, 167–178. doi: 10.1016/j.neuropharm.2018.03.010
- Carhart-Harris, R. L., and Friston, K. J. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
- Carter, C. S., MacDonald, A. W., Ross, L. L., and Stenger, V. A. (2001). Anterior cingulate cortex activity and impaired self-monitoring of performance in patients with schizophrenia: an event-related fMRI study. *Am. J. Psychiatry* 158, 1423–1428. doi: 10.1176/appi.ajp.158.9.1423
- Carvalho, F. (2017). *Excess Dopamine D2R Activation Accounts for PV+ Basket Cell and Learning Alteration* (PhD thesis). Basel University.
- Available online at: https://edoc.unibas.ch/54785/1/Thesis_PhD_Fernando_CarvalhoRodriguesPereira.pdf (accessed May 25, 2022).
- Cascella, N. G., and Sawa, A. (2014). "The claustrum in schizophrenia," In: *The claustrum*. Cambridge: Academic Press, 237–243.
- Cieri, F. (2022). Memory for the future: psychodynamic approach to time and self through the default network. *Front. Hum. Neurosci.* 16, 885315. doi: 10.3389/fnhum.2022.885315
- Cieri, F., Cera, N., Griffo, A., Mantini, D., and Esposito, R. (2020). Editorial: dynamic functioning of resting state networks in physiological and pathological conditions. *Front. Neurosci.* 14, 624401. doi: 10.3389/fnins.2020.624401
- Cieri, F., and Esposito, R. (2018). Neuroaging through the lens of the resting state networks. *Biomed Res. Int.* 2018, 5080981. doi: 10.1155/2018/5080981
- Cieri, F., and Esposito, R. (2019). Psychoanalysis and neuroscience: the bridge between mind and brain. *Front. Psychol.* 10, 1790. doi: 10.3389/fpsyg.2019.01983
- Cieri, F., Zhuang, X., Caldwell, J. Z., and Cordes, D. (2021). Brain entropy during aging through a free energy principle approach. *Front. Hum. Neurosci.* 15, 647513. doi: 10.3389/fnhum.2021.647513
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894
- Damasio, A. (2000). *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*. London: Vintage.
- Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., and Damasio, A. R. (1994). The return of phineas gage: clues about the brain from the skull of a famous patient. *Science* 264, 1102–1105. doi: 10.1126/science.8178168
- Dayan, P., Hinton, G. E., Neal, R. M., and Zemel, R. S. (1995). The helmholtz machine. *Neur. Comput.* 7, 889–904. Available online at: <http://www.gatsby.ucl.ac.uk/dayan/papers/hm95.pdf>
- Donato, F., Chowdhury, A., Lahr, M., and Caroni, P. (2015). Early- and late-born parvalbumin basket cell subpopulations exhibiting distinct regulation and roles in learning. *Neuron* 85, 770–786. doi: 10.1016/j.neuron.2015.01.011
- Dor-Ziderman, Y., Berkovich-Ohana, A., Glicksohn, J., and Goldstein, A. (2013). Mindfulness-induced selflessness: a MEG neurophenomenological study. *Front. Hum. Neurosci.* 7, 582. doi: 10.3389/fnhum.2013.00582
- Doucet, G., Naveau, M., Petit, L., Delcroix, N., Zago, L., Crivello, F., et al. (2011). Brain activity at rest: a multiscale hierarchical functional organization. *J. Neurophysiol.* 105, 2753–2763. doi: 10.1152/jn.00895.2010
- Esposito, R., Cieri, F., Chiacchiarretta, P., Cera, N., Lauriola, M., Di Giannantonio, M., et al. (2018). Modifications in resting state functional anticorrelation between default mode network and dorsal attention network: comparison among young adults, healthy elders and mild cognitive impairment patients. *Brain Imag. Behav.* 12, 127–141. doi: 10.1007/s11682-017-9686-y

- Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., et al. (2008). The maturing architecture of the brain's default network. *Proc. Natl. Acad. Sci. U. S. A.* 105, 4028–4032. doi: 10.1073/pnas.0800376105
- Fodoulou, L., Gschwend, O., Huber, C., Mutel, S., Salazar, R. F., Leone, R., et al. (2020). The claustrum-medial prefrontal cortex network controls attentional set-shifting. *BioRxiv*. doi: 10.1101/2020.10.14.339259
- Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711. doi: 10.1038/nrn2201
- Fox, M. D., Snyder, A. Z., Zacks, J. M., and Raichle, M. E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nat. Neurosci.* 9, 23–25. doi: 10.1038/nn1616
- Fransson, P. (2005). Spontaneous low-frequency BOLD signal fluctuations: an fMRI investigation of the resting-state default mode of brain function hypothesis. *Hum. Brain Mapp.* 26, 15–29. doi: 10.1002/hbm.20113
- Fransson, P., Skjölde, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., et al. (2007). Resting-state networks in the infant brain. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15531–15536. doi: 10.1073/pnas.0704380104
- Freud, S. (1894). “The neuro-psychoses of defence,” in *Standard Edition*, Vol. 3, London: Hogarth Press, 43–61.
- Freud, S. (1895). “Project for a scientific psychology,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. 1, London: Hogarth Press, 283–397.
- Freud, S. (1896a). “Heredity and the aetiology of the neuroses,” in *Standard Edition*, Vol. 3 (London: Hogarth Press), 141–156.
- Freud, S. (1896b). “Draft K,” in *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887–1904*, ed J. M. Masson (Cambridge and London: The Belknap Press of Harvard University Press), 162–169.
- Freud, S. (1898). “Letter to Fliess, 22nd September 1898,” in *The Complete Letters of Sigmund Freud to Wilhelm Fliess, 1887–1904*, ed J. M. Masson (Cambridge and London: The Belknap Press of Harvard University Press), 326–327.
- Freud, S. (1900). “The interpretation of dreams,” in *Standard Edition*, Vol. 4 (London: Hogarth Press).
- Freud, S. (1911). “Psycho-analytic notes on an autobiographical account of a case of paranoia,” in *Standard Edition*, Vol. 12 (London: Hogarth Press), 3–82.
- Freud, S. (1914). “On narcissism: an introduction,” in *Standard Edition*, Vol. 13 (London: Hogarth Press), 67–104.
- Freud, S. (1920). “Beyond the pleasure principle,” in *Standard Edition*, Vol. 18 (London: Hogarth Press), 7–64.
- Freud, S. (1923). “The ego and the Id,” in *Standard Edition*, Vol. 19 (London: Hogarth Press), 1–66.
- Freud, S. (1924a). “Neurosis and psychosis,” in *Standard Edition*, Vol. 19 (London: Hogarth Press), 149–153.
- Freud, S. (1924b). “The loss of reality in neurosis and psychosis,” in *Standard Edition*, Vol. 19 (London: Hogarth Press), 183–187.
- Freud, S. (1925). “A note upon the «mystic writing-pad,” in *Standard Edition*, Vol. 19 (London: Hogarth Press), 227–232.
- Freud, S., and Breuer, J. (1895). “Studies in Hysteria,” in *Standard Edition*. Vol. 2 (London: Hogarth Press), 1955.
- Freud, S. (1915–1916/1963). *Introductory lectures on psycho-analysis: Parts 1 and 2: 1915–1916/The Standard Edition of The Complete Psychological Works of Sigmund Freud*, Sigmund. London: The Hogarth Press. p. 239.
- Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K., Brown, H. R., Siemerkus, J., and Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophr. Res.* 176, 83–94. doi: 10.1016/j.schres.2016.07.014
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053
- Friston, K. J. (2003). Learning and inference in the brain. *Neural Netw.* 16, 1325–52. doi: 10.1016/j.neunet.2003.06.005
- Friston, K. J., and Frith, C. D. (1995). Schizophrenia – a disconnection syndrome. *Clin. Neurosci.* 1995, 89–97.
- Friston, K., Breakspear, M., and Deco, G. (2012). Perception and self-organized instability. *Front. Comput. Neurosci.* 6, 44. doi: 10.3389/fncom.2012.00044
- Galindo, L., Berge, D., Murray, G. K., Man?1, A., Bulbena, A., P?rez, I. V., et al. (2018). Default mode network aberrant connectivity associated with neurological soft signs in schizophrenia patients and unaffected relatives. *Front. Psychiatry* 8, 298. doi: 10.3389/fpsyt.2017.00298
- Garrity, A. G., Pearlson, G. D., McKiernan, K., Lloyd, D., Kiehl, K. A., and Calhoun, V. D. (2007). Aberrant default mode functional connectivity in schizophrenia. *Am. J. Psychiatry* 164, 450–457. doi: 10.1176/ajp.2007.164.3.450
- Genzel, L., Dresler, M., Cornu, M., Jäger, E., Konrad, B., Adamczyk, M., et al. (2015). Medial prefrontal-hippocampal connectivity and motor memory consolidation in depression and schizophrenia. *Biol. Psychiatry* 77, 177–186. doi: 10.1016/j.biopsych.2014.06.004
- Goulden, N., Khusnulnisa, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P., and Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. *Neuroimage* 99, 180–190. doi: 10.1016/j.neuroimage.2014.05.052
- Gregory, R. L. (1968). Perceptual illusions and brain models. *Proc. royal Soc. London Ser. B. Biol. Sci.* 171, 279–296.
- Guo, W., Liu, W., F., Chen, J., Wu, R., Li, L., et al. (2017). Hyperactivity of the default-mode network in first-episode, drug-naïve schizophrenia at rest revealed by family-based case-control and traditional case-control designs. *Medicine* 96, e6223. doi: 10.1097/MD.00000000000006223
- Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694. doi: 10.1038/35094500
- Helmholtz, H. (1866). “Concerning the perceptions in general,” in *Treatise on physiological optics*. Vol. III, 3rd edn (New York, NY: Dover).
- Helmholtz, H. (1882). “On the thermodynamics of chemical processes,” in *Physical Memoirs Selected and Translated from Foreign Sources*, Vol. 1 (London: Taylor and Francis), 43–97.
- Hopkins, J. (2016). Free energy and virtual reality in neuroscience and psychoanalysis: a complexity theory of dreaming and mental disorder. *Front. Psychol.* 7, 922. doi: 10.3389/fpsyg.2016.00922
- Hu, M. L., Zong, X. G., Mann, J. J., Zheng, J.-J., Liao, Y.-L., Li, Z.-C., et al. (2017). A Review of the functional and anatomical default mode network in schizophrenia. *Neurosci. Bull.* 33, 73–84. doi: 10.1007/s12264-016-0090-1
- Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78, 2690. doi: 10.1103/PhysRevLett.78.2690
- Kelly, A. M., Di Martino, A., Uddin, L. Q., Shehzad, Z., Gee, D. G., Reiss, P. T., et al. (2009). Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cereb. Cortex* 19, 640–657. doi: 10.1093/cercor/bhn117
- Kiebel, S. J., Daunizeau, J., and Friston, K. J. A. (2008). Hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4, e1000209. doi: 10.1371/journal.pcbi.1000209
- Koubeissi, M. Z., Bartolomei, F., Beltagy, A., and Picard, F. (2014). Electrical stimulation of a small brain area reversibly disrupts consciousness. *Epilepsy Behav.* 37, 32–35. doi: 10.1016/j.yebeh.2014.05.027
- Lacan, J. (1946). “Presentation on psychical causality,” in *Écrit. The First Complete English Edition*, ed B. Fink (New York, NY, London: W.W. Norton), 123–158.
- Lafargue, G. (2010). “Dernières découvertes,” in *Un cerveau jamais au repos*, Pour la science, n° 393, ed M. Raichle (Paris: Editions Pour la science), 47.
- Laplanche, J., and Pontalis, J. B. (1988). Free energy/ bound energy,” in *The Language of Psycho-Analysis*, ed J. Laplanche and J. B. Pontalis (London: Karnac Books), 171–173.
- Lee, H., Lee, D. K., Park, K., Kim, C. E., and Ryu, S. (2019). Default mode network connectivity is associated with long-term clinical outcome in patients with schizophrenia. *NeuroImage Clin.* 22, 101805. doi: 10.1016/j.nicl.2019.101805
- Liu, H., Kaneko, Y., Ouyang, X., Li, L., Hao, Y., Chen, E. Y., et al. (2012). Schizophrenic patients and their unaffected siblings share increased resting-state connectivity in the task-negative network but not its anticorrelated task-positive network. *Schizophr. Bull.* 38, 285–294. doi: 10.1093/schbul/sbq074
- Mallikarjun, P. K., Lalouis, P. A., Dunne, T. F., Heinze, K., Reniers, R. L., Broome, M. R., et al. (2018). Aberrant salience network functional connectivity in auditory verbal hallucinations: a first episode psychosis sample. *Transl. Psychiatry* 8, 96. doi: 10.1038/s41398-018-0118-6

- Molnar-Szakacs, I., and Uddin, L. Q. (2013). Self-processing and the default mode network: interactions with the mirror neuron system. *Front. Hum. Neurosci.* 7, 571–571. doi: 10.3389/fnhum.2013.00571
- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00198477
- Ortega, P. A., and Braun, D. A. (2010). A minimum relative entropy principle for learning and acting. *J. Artif. Intell. Res.* 38, 475–511. doi: 10.1613/jair.3062
- Oswald, W. (1891). Studien zur Energetik I. *Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig.* 43, 271–288.
- Park, J., and Moghaddam, B. (2017). Impact of anxiety on prefrontal cortex encoding of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16, 331–348. doi: 10.1006/nimg.2002.1087
- Prigogine, I. (1967). *Introduction to Thermodynamics of Irreversible Processes*, 3rd Edn (New York, NY: John Wiley and Sons).
- Prigogine, I., and Stengers, I. (1984). *Order Out of Chaos*. New York, NY: Bantam.
- Rabeyron, T. (2021). Beyond the death drive: entropy and free energy. *Int. J. Psychoanal.* 102, 878–905. doi: 10.1080/00207578.2021.1932514
- Rabeyron, T. (2022). Psychoanalytic psychotherapies and the free energy principle. *Front. Hum. Neurosci.* 16, 929940. doi: 10.3389/fnhum.2022.929940
- Rabeyron, T., and Massicotte, C. (2020). Entropy, free energy, and symbolization: free association at the intersection of psychoanalysis and neuroscience. *Front. Psychol.* 11, 366. doi: 10.3389/fpsyg.2020.00366
- Raichle, M. E. (2001). Bold insights. *Nature.* 412, 128–130. doi: 10.1038/35084300
- Raichle, M. E. (2006). Neuroscience. The brain's dark energy. *Science* 314, 1249–1250. doi: 10.1126/science.1134405
- Raichle, M. E. (2010). Two views of brain function. *Trends Cogn. Sci.* 14, 180–190. doi: 10.1016/j.tics.2010.01.008
- Raichle, M. E. (2015). The brain's default mode network. *Annu. Rev. Neurosci.* 38, 433–447. doi: 10.1146/annurev-neuro-071013-014030
- Raichle, M. E., and Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proc. Natl. Acad. Sci.* 99, 10237–10239. doi: 10.1073/pnas.172399499
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci.* 98, 676–682. doi: 10.1073/pnas.98.2.676
- Ramstead, M. J. D., Badcock, P. J., and Friston, K. J. (2018). Answering Schrödinger's question: a free-energy formulation. *Phys. Life Rev.* 24, 1–16. doi: 10.1016/j.plrev.2017.09.001
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Rapaport, D. (1958). *The Structure of Psychoanalytic Theory (a Systematizing attempt) Psychology, A Study of a Science*, 3. New York, NY: S. Koch.
- Reichenbach, J. R., Schlösser, R. G., Gaser, C., Sauer, H., and Nenadic, I. (2012). Default mode network activity in schizophrenia studied at resting state using probabilistic ICA. *Schizophr. Res.* 138, 143–149. doi: 10.1016/j.schres.2012.01.036
- Salman, M. S., Vergara, V. M., Damaraju, E., and Calhoun, V. D. (2019). Decreased cross-domain mutual information in schizophrenia from dynamic connectivity states. *Front. Neurosci.* 13, 873. doi: 10.3389/fnins.2019.00873
- Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *NeuroImage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1
- Schreber, D. P. (2000). *Memoirs of My Nervous Illness*, eds I. Macalpine and R. A. Hunter (New York, NY: New York Review Books).
- Sérieux, P., and Capgras, J. (1982). *Les folies raisonnantes, le délire d'interprétation*. Marseille: Laffite Reprints.
- Shannon, B. J., Dosenbach, R. A., Su, Y., Vlassenko, A. G., Larson-Prior, L. J., Nolan, T. S., et al. (2013). Morning-evening variation in human brain metabolism and memory circuits. *J. Neurophysiol.* 109, 1444–1456. doi: 10.1152/jn.00651.2012
- Shim, G., Oh, J. S., Jung, W. H., Jang, J. H., Choi, C.-H., Kim, E., et al. (2010). Altered resting-state connectivity in subjects at ultra-high risk for psychosis: an fMRI study. *Behav. Brain Funct.* 6, 58–58. doi: 10.1186/1744-9081-6-58
- Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F., M., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci.* 9, 648–663. doi: 10.1162/jocn.1997.9.5.648
- Skudlarski, P., Jagannathan, K., Anderson, K., Stevens, M. C., Calhoun, V. D., Skudlarska, B. A., et al. (2010). Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. *Biol. Psychiatry* 68, 61–69. doi: 10.1016/j.biopsych.2010.03.035
- Sokoloff, L., Mangold, R., Wechsler, R. L., Kenney, C., and Kety, S. (1955). The effect of mental arithmetic on cerebral circulation and metabolism. *J. Clin. Investig.* 34, 1101–1108. doi: 10.1172/JCI103159
- Sokunbi, M. O., Gradin, V. B., Waiter, G. D., Cameron, G. G., Ahearn, T. S., Murray, A. D., et al. (2014). Nonlinear complexity analysis of brain fMRI signals in schizophrenia. *PLoS ONE* 9, e95146. doi: 10.1371/journal.pone.0095146
- Solms, M. (2013). The conscious if. [and response to commentaries]. *Neuropsychanalysis* 15, 5–85. doi: 10.1080/15294145.2013.10773711
- Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9, 2714. doi: 10.3389/fpsyg.2018.02714
- Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious Stud.* 25, 202–238.
- Spreng, R. N., Mar, R. A., and Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510. doi: 10.1162/jocn.2008.21029
- Sridharan, D., Levitin, D. J., and Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl. Acad. Sci. U. S. A.* 105, 12569–12574. doi: 10.1073/pnas.0800005105
- Svoboda, E., McKinnon, M. C., and Levine, B. (2006). The functional neuroanatomy of autobiographical memory: a meta-analysis. *Neuropsychologia* 44, 2189–2208. doi: 10.1016/j.neuropsychologia.2006.05.023
- Teves, D., Videen, T. O., Cryer, P. E., and Powers, W. J. (2004). Activation of human medial prefrontal cortex during autonomic responses to hypoglycemia. *Proc. Natl. Acad. Sci.* 101, 6217–6221. doi: 10.1073/pnas.0307048101
- Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242. doi: 10.2307/25470707
- Tran The, J., Ansermet, J. P., Magistretti, P., and Ansermet, F. (2020). From the principle of inertia to the death drive: the influence of the second law of thermodynamics on the Freudian theory of the psychical apparatus. *Front. Psychol.* 11, 325. doi: 10.3389/fpsyg.2020.00325
- Tran The, J., Magistretti, P., and Ansermet, F. (2018). The epistemological foundations of Freud's energetics model. *Front. Psychol.* 9, 1861. doi: 10.3389/fpsyg.2018.01861
- Tran The, J., Magistretti, P., and Ansermet, F. (2021). Interoception disorder and insular cortex abnormalities in schizophrenia: a new perspective between psychoanalysis and neuroscience. *Front. Psychol.* 12, 628355. doi: 10.3389/fpsyg.2021.628355
- Turnbull, O. H., and Solms, M. (2007). Awareness, desire, and false beliefs: Freud in the light of modern neuropsychology. *Cortex* 43, 1083–1090. doi: 10.1016/S0010-9452(08)70706-8
- Vincent, J. L., Snyder, A. Z., Fox, M. D., Shannon, B. J., Andrews, J. R., Raichle, M. E., et al. (2006). Coherent spontaneous activity identifies a hippocampal-parietal memory network. *J. Neurophysiol.* 96, 3517–3531. doi: 10.1152/jn.00048.2006
- Whitfield-Gabrieli, S., and Ford, J. M. (2012). Default mode network activity and connectivity in psychopathology. *Annu. Rev. Clin. Psychol.* 8, 49–76. doi: 10.1146/annurev-clinpsy-032511-143049
- Whitfield-Gabrieli, S., Thermenos, H. W., Milanovic, S., Tsuang, M. T., Faraone, S. V., McCarley, R. W., et al. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proc. Natl. Acad. Sci.* 106, 1279–1284. doi: 10.1073/pnas.0809141106
- Yin, B., Terhune, D. B., Smythies, J., and Meck, W. H. (2016). Claustrum, consciousness, and time perception. *Curr. Opin. Behav. Sci.* 8, 258–267. doi: 10.1016/j.cobeha.2016.02.032
- Zhao, Z., Li, X., Feng, G., Shen, Z., Li, S., Xu, Y., et al. (2018). Altered effective connectivity in the default network of the brains of first-episode, drug-naïve schizophrenia patients with auditory verbal hallucinations. *Front. Hum. Neurosci.* 12, 456–456. doi: 10.3389/fnhum.2018.00456
- Zhou, Y., Liang, M., Tian, L., Wang, K., Hao, Y., Liu, H., et al. (2007). Functional disintegration in paranoid schizophrenia using resting-state fMRI. *Schizophr. Res.* 97, 194–205. doi: 10.1016/j.schres.2007.05.029



OPEN ACCESS

EDITED BY

Andy Wai Kan Yeung,
University of Hong Kong, China

REVIEWED BY

Jessica Tran The,
INSERM U1077 Neuropsychologie et
Imagerie de la Mémoire
Humaine, France
Anna Buchheim,
University of Innsbruck, Austria

*CORRESPONDENCE

Filippo Cieri
filippocieri@gmail.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 27 August 2022

ACCEPTED 25 November 2022

PUBLISHED 20 December 2022

CITATION

Cera N, Monteiro J, Esposito R, Di
Francesco G, Cordes D, Caldwell JZK
and Cieri F (2022) Neural correlates of
psychodynamic and
non-psychodynamic therapies in
different clinical populations through
fMRI: A meta-analysis and systematic
review.
Front. Hum. Neurosci. 16:1029256.
doi: 10.3389/fnhum.2022.1029256

COPYRIGHT

© 2022 Cera, Monteiro, Esposito, Di
Francesco, Cordes, Caldwell and Cieri.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Neural correlates of psychodynamic and non-psychodynamic therapies in different clinical populations through fMRI: A meta-analysis and systematic review

Nicoletta Cera^{1,2†}, Jessica Monteiro¹, Roberto Esposito³,
Giulia Di Francesco⁴, Dietmar Cordes^{5,6,7},
Jessica Z. K. Caldwell⁵ and Filippo Cieri^{5*†}

¹Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal, ²CIBIT-Coimbra Institute for Biomedical Imaging and Translational Research, Coimbra, Portugal, ³Department of Radiology, Area Vasta 1/ASUR Marche, Pesaro, Italy, ⁴Hematology Unit, Pescara Hospital, Pescara, Italy, ⁵Department of Neurology, Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, United States, ⁶Department of Brain Health, University of Nevada, Las Vegas, NV, United States, ⁷Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, United States

Background: The COVID-19 pandemic has exacerbated the ongoing crisis in psychiatric and psychological care, contributing to what we have identified as *a new psychological and psychiatric pandemic*. Psychotherapy is an effective method for easing the psychological suffering experienced also by the various impacts of COVID-19. This treatment can be examined from a neurological perspective, through the application of brain imaging techniques. Specifically, the meta-analysis of imaging studies can aid in expanding researchers' understanding of the many beneficial applications of psychotherapy.

Objectives: We examined the functional brain changes accompanying different mental disorders with functional Magnetic Resonance Imaging (fMRI), through a meta-analysis, and systematic review in order to better understand the general neural mechanism involved in psychotherapy and the potential neural difference between psychodynamic and non-psychodynamic approaches.

Data sources: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were employed for our systematic review and meta-analysis. We conducted a computer-based literature search, following the Population, Intervention, Comparison and Outcomes (PICO) approach, to retrieve all published articles in English regarding the above-described topics from PubMed (MEDLINE), Scopus, and Web of Science.

Study eligibility criteria, participants, and interventions: We combined terms related to psychotherapy and fMRI: ("psychotherapy" [All Fields] OR "psychotherapy" [MeSH Terms] OR "psychotherapy" [All Fields] OR "psychotherapies" [All Fields] OR "psychotherapy s" [All Fields]) AND ("magnetic resonance imaging" [MeSH Terms]) OR ("magnetic" [All Fields] AND

"resonance"[All Fields] AND "imaging"[All Fields]) OR ("magnetic resonance imaging"[All Fields] OR "fmri"[All Fields]). We considered (1) whole brain fMRI studies; (2) studies in which participants have been involved in a clinical trial with psychotherapy sessions, with pre/post fMRI; (3) fMRI results presented in coordinate-based (x, y, and z) in MNI or Talairach space; (4) presence of neuropsychiatric patients. The exclusion criteria were: (1) systematic review or meta-analysis; (2) behavioral study; (3) single-case MRI or fMRI study; and (4) other imaging techniques (i.e., PET, SPECT) or EEG.

Results: After duplicates removal and assessment of the content of each published study, we included 38 sources. The map including all studies that assessed longitudinal differences in brain activity showed two homogeneous clusters in the left inferior frontal gyrus, and caudally involving the anterior insular cortex ($p < 0.0001$, corr.). Similarly, studies that assessed psychotherapy-related longitudinal changes using emotional or cognitive tasks (TASK map) showed a left-sided homogeneity in the anterior insula ($p < 0.000$) extending to Broca's area of the inferior frontal gyrus ($p < 0.0001$) and the superior frontal gyrus ($p < 0.0001$). Studies that applied psychodynamic psychotherapy showed Family-Wise Error (FWE) cluster-corrected ($p < 0.05$) homogeneity values in the right superior and inferior frontal gyri, with a small cluster in the putamen. No FWE-corrected homogeneity foci were observed for Mindful-based and cognitive behavioral therapy psychotherapy. In both pre- and post-therapy results, studies showed two bilateral clusters in the dorsal anterior insulae ($p = 0.00001$ and $p = 0.00003$, respectively) and involvement of the medial superior frontal gyrus ($p = 0.0002$).

Limitations: Subjective experiences, such as an individual's response to therapy, are intrinsically challenging to quantify as objective, factual realities. Brain changes observed both pre- and post-therapy could be related to other factors, not necessary to the specific treatment received. Therapeutic modalities and study designs are generally heterogeneous. Differences exist in sample characteristics, such as the specificity of the disorder and number and duration of sessions. Moreover, the sample size is relatively small, particularly due to the paucity of studies in this field and the little contribution of PDT.

Conclusions and implications of key findings: All psychological interventions seem to influence the brain from a functional point of view, showing their efficacy from a neurological perspective. Frontal, prefrontal regions, insular cortex, superior and inferior frontal gyrus, and putamen seem involved in these neural changes, with the psychodynamic more linked to the latter three regions.

KEYWORDS

neural correlate, psychotherapy, psychoanalysis, meta-analysis, fMRI, psychodynamic neuroscience

Introduction

Mental disorders represent a significant public health concern, producing an enormous economic burden for society and great suffering for patients as well as their families and communities. After almost 3 years of the COVID-19 pandemic,

the situation has worsened, creating increased urgency to strengthen mental health systems in most countries (COVID-19 Mental Disorders Collaborators, 2021).

Although the biopsychosocial model is shared by the World Health Organization's International Classification of Functioning (WHO ICF), the previous biomedical model of

mental disorders is nowadays anachronistic but still pervasive, with supporters among clinicians and researchers, thereby threatening the understanding of highly complex phenomena, such as the human mental suffering. In addition to a more complex etiopathogenesis of mental disorders supported by the biopsychosocial model, compared to the biomedical model, a more complex approach to therapy also exists. Among the major criticisms of the biomedical model, especially related to mental disorders, its reductive view of etiopathogenesis and approach to therapy are the most important concerns.

Pre-COVID studies have shown that at least one person out of two (50% of the high-income countries' population) will be diagnosed with a psychiatric disorder in their life (Wittchen et al., 2011), most commonly anxiety and depressive disorders. Antidepressants are currently among the most frequently prescribed medications worldwide, being taken by more than 10% of the general population annually in high-income countries (Jorm et al., 2017; Furukawa et al., 2021). More patients every year are on longer-term anti-depressant treatment. In the US, even before the COVID-19 pandemic, we observed an increase of almost 30% of patients who had been using antidepressants for more than 5 years, from 13% in 1996, to 44% in 2015 (Jorm et al., 2017; Luo et al., 2020). Still in the US, we observed an increase of 21% in the number of antidepressant prescriptions in February and March 2020, the highest point since the declaration of the COVID-19 pandemic (America's State of Mind Report, 2020).

This paper is neither about the efficacy of pharmacotherapy, nor the efficacy of psychotherapy, or the comparison between psychotherapy and pharmacotherapy, but a brief introduction can be helpful to establish the urgency of this topic described by our current work and the necessity of studies able to explore the effects of psychological therapies, from a behavioral and neurological point of view, especially in a post-pandemic era, in which society is experiencing a *new psychological and psychiatric pandemic*.

Pharmacotherapy and psychotherapy

Taking into consideration only the most widespread source of mental suffering, with its pharmacological treatment, according to the Food and Drug Administration (FDA; Turner et al., 2008), the efficacy of antidepressant drugs for depression is quite small, with an effect size of 0.26 for Fluoxetine and Sertraline, 0.24 Citalopram, 0.31 for Escitalopram, 0.30 for Duloxetine, and with an overall effect size of FDA approved antidepressant drugs between 1987 and 2004 of 0.31. Other more recent studies have shown that antidepressant medications achieve effect sizes of between 0.24 (tricyclics) and 0.31 in case of SSRIs (Kirsch et al., 2008). When Kirsch et al. (2008) have used published and unpublished Food

and Drug Administration (FDA) registration trials, to assess antidepressant efficacy, they found that antidepressants were not clinically significant for mild, moderate, and severe depression, with a mean drug-placebo difference of only 1.80 points on the Hamilton Depression Rating Scale. Some studies suggest there are medications less effective than a placebo (clomipramine; Cipriani et al., 2018). A more recent study (Almohammed et al., 2022) has considered changes in the quality of life over 2 years in Americans with depression who took antidepressants (any type) vs. the changes reported by those with the same diagnosis who did not use antidepressants. The study found no significant difference in the quality of life of this population.

A recent review has revealed a lack of randomized double-blind placebo-controlled trials for anxiety disorders and few studies comparing novel treatments to existing anxiolytic agents, concluding that although some randomized controlled trials for novel agents exist, these trials have largely been negative (Garakani et al., 2020).

Furukawa et al. (2021) have recently published a systematic review and meta-analysis of randomized controlled trials (RCTs) in which adult patients with major depressive disorder were randomized to acute treatment with a psychotherapy (PSY), a protocolized antidepressant pharmacotherapy (PHA), their combination (COM), standard treatment in primary or secondary care, or pill placebo, and were then followed up through a maintenance phase. According to this authoritative paper, psychotherapy shows more effectiveness than pharmacotherapy, both if these treatments are continued into the maintenance phase (PSY → PSY vs. PHA → PHA: OR = 1.53, 95% CI: 1.00–2.35) and if they were followed by discretionary treatment (PSY → naturalistic vs. PHA → naturalistic: OR = 1.66, 95% CI: 1.13–2.44). The same applied to PSY when compared with standard therapy through the acute and maintenance phases [PSY → PSY vs. standard treatment in primary or secondary care (STD): OR = 1.76, 95% CI: 0.97–3.21; PSY → nat vs. STD: OR = 1.83, 95% CI: 1.20–2.78]. In other words, PSY (and combination of PSY and PHA) has more enduring effects than PHA. Therefore, guidelines on the treatment choice for depression may need to be updated accordingly (Furukawa et al., 2021).

Another recent systematic review (Wakefield et al., 2021) confirmed the efficacy of psychotherapy, showing large pre-post treatment effect sizes for depression [$d = 0.87$, 95% CI (0.78–0.96), $p < 0.0001$] and anxiety [$d = 0.88$, 95% CI (0.79–0.97), $p < 0.0001$]. In the comparison between cognitive behavioral (CBT) and psychodynamic psychotherapy (PDT) studies and meta-analysis have shown similar efficacy (Leichsenring and Steinert, 2017; Steinert et al., 2017), although some authors point out that in the case of PDT the effects last longer—and even increase—after the end of the treatment, with an effect size of between 0.78 and 1.46, even for diluted and truncated forms of psychoanalytic therapy (Shedler, 2010; Solms, 2018). A quasi-experimental comparison found psychoanalysis but not

long-term PDT to be superior to CBT on measures of depression at 3-year follow-up (Huber et al., 2012). However, some authors have found a worse performance of PDT compared to CBT (Barber et al., 2021).

When we compare PHA and PDT, some studies fail to identify differential effects (Salminen et al., 2008; Barber et al., 2012; Bloch et al., 2012; Zilcha-Mano et al., 2014). On the other hand, a recent meta-review of 61 meta-analyses covering 21 psychiatric disorders containing 852 trials and 137,126 subjects yielded larger effect sizes for PDT (0.58; 95% CI: 0.42–0.76) than PHA (0.40; 95% CI: 0.28–0.52) studies (Huhn et al., 2014).

All these studies clearly show that psychotherapy (PDT, CBT, and other psychological approaches) works, usually more effectively than PHA alone, and it works by modifying patients' symptoms, thinking patterns, beliefs, attitudes, emotional states, and behaviors, in the most widespread mental disorders, showing a good effect size, in general and compared to pharmacotherapy. Based on these results, it should be a priority of the clinical and research community to understand not if PSY works, but how this form of treatment is able to act on a neurological and neuropsychological level. Without claiming to be exhaustive, the current study uses meta-analytic and systematic review approaches to explore functional brain changes through functional Magnetic Resonance Imaging (fMRI) among different mental disorders, trying to understand the general neural mechanism involved in psychotherapy and potential differences between changes associated with PDT and non-PDT.

Neural effects

As Marek et al. (2009) have recently pointed out, most brain-wide association studies (typically based on a sample size of 25 subjects) have shown inadequacy. These authors have used a meta-analytic approach on three of the largest neuroimaging datasets currently available—with a total sample size of around 50,000 individuals, revealing how the usual brain wide association studies are not appropriate in capturing inter-individual differences in brain structure or function and complex cognitive or mental health phenotypes (Marek et al., 2009). Therefore, together with big data, a meta-analytic approach can enlarge the sample size, increase the statistical power, and give a more accurate idea about general brain changes. From a psychological (especially psychodynamic) point of view, we also note that subjectivity should be taken into consideration, given this form of therapy is highly individualized, with the *cure* passing through the therapist; thus, the subjectivity of the therapist, the patient, and their relationship, should be taken in consideration. However, since meta-analyses are not based on individuality, but on mean values, this search for objectivity has historically kept some

psychoanalysts distant or skeptical toward research in general and neuroscientific research in particular.

If subjectivity is given more consideration by clinicians, therapists, and scientists involved in individual-differences level research, using a meta-analytic approach, a compromise between quantity (sample size) and quality (subjectivity) might be more reasonably reached. Such a compromise would meet the need to reinforce an already existing bridge between neuroscience and psychoanalysis, a process yearned for by psychoanalysts and neuroscientists alike over the last 20 years, and started by Freud more than a century ago (Kandel, 1999, 2016; Kaplan-Solms and Solms, 2000; Solms and Turnbull, 2002; Carhart-Harris and Friston, 2010; Boeker et al., 2013; de Greck et al., 2013; Scalabrini et al., 2018; Solms, 2018, 2021; Cieri and Esposito, 2019; Cieri et al., 2021; Northoff and Scalabrini, 2021; Rabeyron, 2021, 2022; Cieri, 2022).

Effects of psychological therapies, similarly to pharmacotherapy, are capable of visualization through brain imaging methods (Kandel, 1999). Meta-analytic approaches of psychotherapies can play a fundamental role understanding their neural, together with their psychological effects. This approach uses brain imaging methods to explore potential structural or functional effects of the talking cure on the brain, finding similar dysfunction in limbic structures, amygdala, hippocampus, frontal cortex, cingulate cortex, and basal ganglia (Sözeri-Varma and Karadağ, 2012).

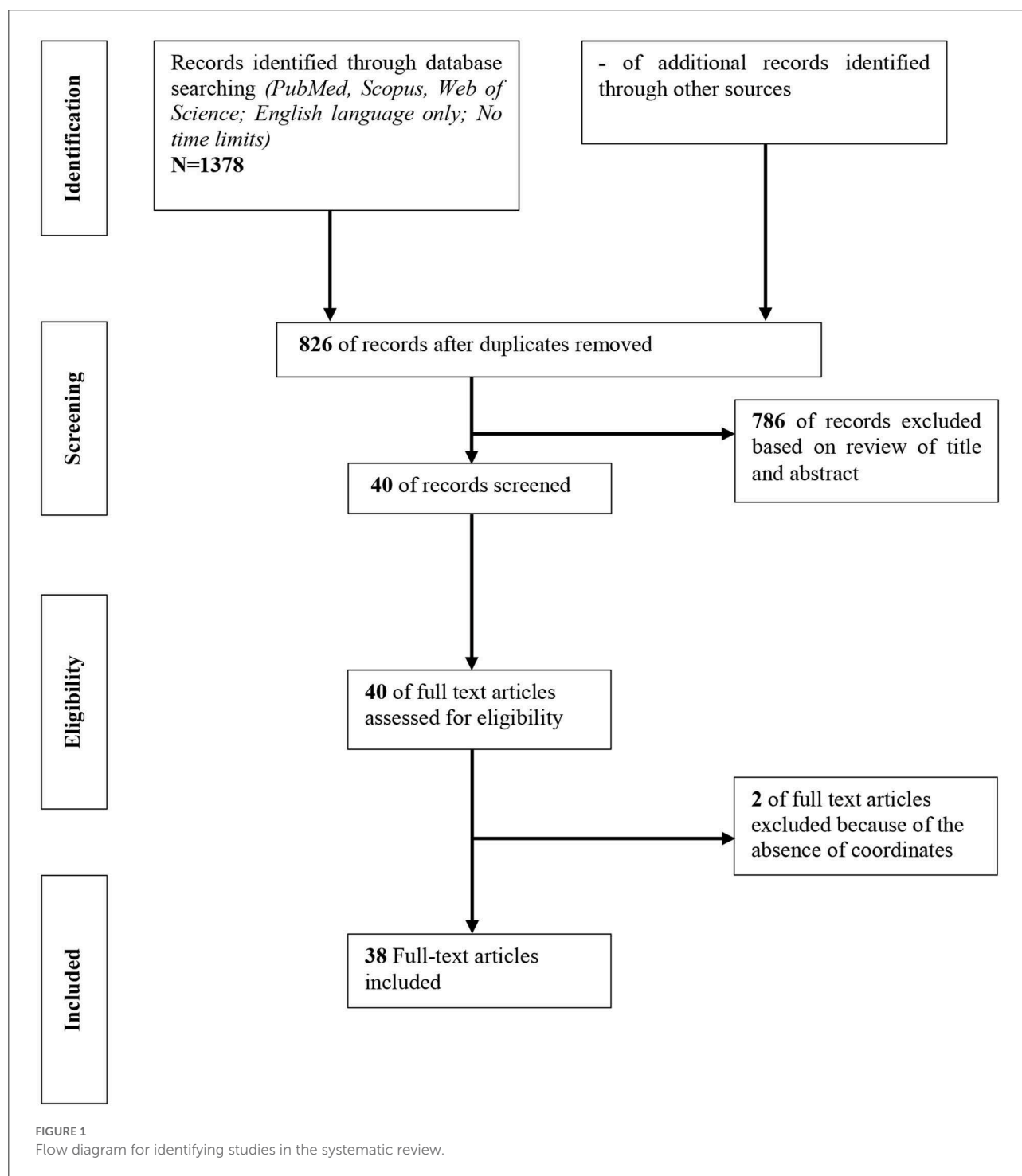
Abbass et al. (2012) described a meta-analysis of brain imaging studies from 11 sources analyzing any form of PDT treatment. The sample was composed of 2 randomized controlled trials, 5 controlled trials and 4 case series. The patient's cohort was affected by depression (atypical and typical), borderline personality disorder, panic disorder and somatoform disorder, investigated by a variety of neuroimaging methods to examine regional metabolic activity and synaptic neurotransmission before and after treatment. These authors found a general normalization of synaptic or metabolic activity in limbic, mid-brain and prefrontal regions, occurring in association with improved clinical outcomes. Patterns of neural activity or neurophysiological infrastructure in regions of the dorsolateral prefrontal cortex (DLPFC), orbital frontal cortex (OFC), anterior cingulate cortex (ACC), and amygdala were found to vary between patients and healthy controls before the psychotherapy, while after treatment, the patterns seen in patients resembled those of the controls.

Messina et al. (2013) conducted a similar meta-analytic study including 16 sources, regardless of the specific psychotherapy approach used. Different diagnosis and methods were analyzed: depression, post-traumatic stress disorder, and panic disorder, investigated both with resting state and task-related activation. They have also considered phobic patients through exposure-related activation method. Anxiety and depression studies showed consistent results for changes in the dorsomedial prefrontal cortex (DMPFC) and

in the posterior cingulate cortex/precuneus (PCC/Prc). Some changes were also described at the level of temporal lobes, both in anxious/depressed and phobic patients.

We can find analogous results in similar regions coming from the systematic review by Franklin et al. (2016). They analyzed brain changes, taking in consideration 10 neuroimaging studies associated with cognitive behavioral

therapy (CBT) of depression. This specific form of psychological treatment was mostly correlated with changes in the ACC, PCC, VMPFC/OFC, and amygdala/hippocampus. As the authors suggest, CBT appeared to decrease the resting state activity in the dorsal ACC. Researchers involved in this study suggest that this form of treatment can develop an increased capacity for “top-down” emotion



regulation, which is employed when skills taught in CBT are engaged.

Another systematic review was conducted by [Gotink et al. \(2016\)](#) taking into consideration both the structural and functional neuronal in stress-reducing effects of the 8-week Mindfulness Based Stress Reduction (MBSR) and Mindfulness Based Cognitive Therapy (MBCT) program. They considered 21 fMRI studies, showing that functional and structural changes in the prefrontal cortex (PFC), cingulate cortex, insula, and hippocampus are similar to changes described in studies on traditional meditation practice. In addition, MBSR led to changes in the amygdala consistent with improved emotion regulation.

[Sankar et al. \(2018\)](#) focused their meta-analytic investigation on psychotherapy of major depressive disorder (MDD), measuring neural function and metabolism using functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), Single-photon Emission Computerized Tomography (SPECT) and Magnetic Resonance Spectroscopy (MRS). A significant group by time effect was found in left rostral ACC, in which patients showed increased activity following psychotherapy while healthy controls showed a decrease at follow up. Longitudinal treatment effects revealed reduced left precentral cortical activity in MDD patients. Findings could be indicative of improvements in emotion responsivity that may be achieved following a psychological treatment, as suggested by the authors of this study.

A more recent study comes from [Thorsen et al. \(2018\)](#) who conducted a systematic review and performed a meta-analysis (Seed-based d-Mapping) of 25 whole-brain neuroimaging studies using fMRI or PET comparing brain activation of Obsessive Compulsive Disorder (OCD) patients and healthy controls during presentation of emotionally-valenced vs. neutral stimuli. OCD patients show increased emotional processing-related activation in limbic, frontal, and temporal regions, compared to healthy controls. We can observe also here similar results in similar regions described in the above mentioned studies. Particularly, patients showed increased activation in the bilateral amygdala, right putamen, OFC extending into the ACC and VMPFC, middle temporal, and left inferior occipital cortices during emotional processing.

Aims of the study

The objective of this meta-analysis and systematic review was to describe the neural correlates of psychological treatments. The distinctive element of our approach derives from the fact that we sought to systematically review the functional neural effects of psychotherapy, through the fMRI, both through the resting-state (rs-fMRI), and task-fMRI approaches, trying to limit the heterogeneity of imaging's tools used. We eliminated from our study structural methods (volumetric,

cortical thickness, white matters etc.) and other brain imaging approaches (PET, SPECT, MRS, etc.). We considered an initial sample of 1,378 studies, reduced to 38 with a total sample size of 1,688 subjects ([Figure 1](#)). Another specificity of our approach is from a demographic and diagnostic point of views. We used the studies on adulthood, with an age range 18–65 years old, with two time points: before and after treatment; analyzing functional neural changes in major depressive disorder (MDD; 11 studies; 366 subjects), panic disorder (PD; 1 study; 27 subjects), somatoform disorder (SD; 2 studies; 120 subjects), social anxiety disorder (SAD; 5 studies; 169 subjects), generalized anxiety disorder (GAD; 2 studies; 57 subjects), post-traumatic stress disorder (PTSD; 7 studies; 412 subjects), obsessive compulsive disorder (OCD; 4 studies; 191 subjects), attention deficit hyperactivity disorder (ADHD; 1 study; 40 subjects), anhedonia (1 study; 73 subjects) and schizophrenia (SZ; 4 studies; 156 subjects). Moreover, from a psychotherapeutic perspective, we tried to outline as far as possible—conditioned by statistical limits deepened in the course of our work—the difference between psychodynamic (PDT; 4 studies) and non-PDT approaches (34 studies).

Methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines has been used for our systematic review and meta-analysis ([Moher et al., 2009](#)).

In particular, our research question was specifically oriented toward functional neural changes related to psychotherapy intervention with different approaches (O) in psychologically healthy controls and psychiatric patients (P). The present review and meta-analysis was limited to longitudinal fMRI studies (I) with tasks and resting state (C). We defined the search terms based on the above mentioned PICO question combined with the Boolean operators “AND” and “OR”, according to the method previously used in other previously published systematic reviews ([Cera et al., 2021](#); [Vargas-Cáceres et al., 2021](#)). We conducted a computer-based literature search follow the PICO approach combining terms related to psychotherapy and fMRI: (“psychotherapy” [All Fields] OR “psychotherapy” [MeSH Terms] OR “psychotherapy” [All Fields] OR “psychotherapies” [All Fields] OR “psychotherapy s” [All Fields]) AND (“magnetic resonance imaging” [MeSH Terms]) OR (“magnetic” [All Fields] AND “resonance” [All Fields] AND “imaging” [All Fields]) OR (“magnetic resonance imaging” [All Fields] OR “fmri” [All Fields]). We conducted the computer-based literature search to retrieve all the published articles in English regarding the above-described topics. We conducted our search in the three principal databases: PubMed (MEDLINE), Scopus and Web of Science.

Importantly, we identify fMRI studies on the basis of the following inclusion criteria: (1) whole brain fMRI studies; (2)

Studies in which participants have been involved in a clinical trial with psychotherapy sessions; (3) fMRI results presented in coordinate-based (x, y, and z) in MNI or Talairach space; (4) presence of neuropsychiatric patients. The exclusion criteria were: (1) systematic review or meta-analysis; (2) behavioral study; (3) single-case MRI or fMRI study; (4) other imaging techniques (i.e., PET and SPECT) or EEG; (5) no coordinates.

The studies, which did not meet the above-mentioned criteria for the meta-analysis, will be included in our systematic review.

All of the included studies were screened to identify additional relevant bibliographic items. Similarly, the narrative, systematic reviews and meta-analyses were retrieved, and we screened them to find previous relevant articles in the reference lists.

After the duplicate removal, the title and abstracts were manually screened to determine if they fulfilled the inclusion and/or exclusion criteria. After the retrieval of potentially relevant studies, we read the full texts to confirm their eligibility.

Two authors conducted independently the literature search, screening, and methodological evaluation. The consensus about the different stages was reached between the two authors discussing the results and the articles retrieved. If a consensus was not reached, a third opinion was obtained.

To assess the quality of the studies included in the present systematic review, we applied the “NOS-scale” (Wells et al., 2022). Information has been extracted from each included study, following the above-mentioned guidelines. In particular, we extracted the characteristics of the participants, including the exclusion and inclusion criteria.

ALE meta-analysis of the brain clusters resulting from the included fMRI studies

To assess the presence of a set of brain regions involved in the longitudinal above-mentioned studies, the brain coordinates reported in each included article were used for an ALE (activation likelihood estimation) meta-analysis. GingerAle 3.02 (Turkeltaub et al., 2012-<https://www.brainmap.org/ale/>) has been used to calculate the ALE meta-analysis. Indeed, GingerAle generates modeled maps of brain activations, by combining the probabilities of all brain activation foci for each voxel, as reported in the studies. Thus, the resulting maps are combined to obtain a voxel-wise ALE score. The ALE scores represent the convergence of results derived from the studies into a functional anatomical localization in the brain. The scores obtained are then compared with an empirical null distribution that represents a between—studies causal association (Eickhoff et al., 2012). Moreover, all coordinates following the stereotaxic space of the Montreal

Neurological Institute-Hospital (MNI) have been transformed into the standardized 3D space of Talairach and Tournoux (1988) using Bioimage Suite (<https://bioimagesuiteweb.github.io/webapp/mni2tal.html>). To understand the role played by different brain regions in psychotherapy, we conducted three different ALE meta-analyses studies, with the cluster-level FWE correction ($p < 0.01$), where possible, or uncorrected statistical threshold with $p < 0.005$. According to Radua and Mataix-Cols (2012), a p threshold of 0.005 is reasonable. Moreover, to disentangle the contribution of the task-based and resting state studies, and to facilitate the interpretation of the results, a conjunction analysis has been performed using GingerALE, using a statistical threshold of $p < 0.05$ uncorrected. We used Mango 4.1 (<http://ric.uthscsa.edu/mango/mango.html>; Lancaster et al., 2010), which allows the visualization of results obtained by navigating between the volumes of the image of an MRI template in the Talairach stereotaxic space (1988) with $2 \times 2 \times 2$ mm resolution (<https://www.brainmap.org/ale/>). Due to no contrast $T2 > T1$ (or interaction effect group $\times$ time) results reported in 2 studies, 36 studies have been included in the present meta-analysis.

Results

After duplicates removal and the assessment of the content of each published study, we included 38 sources. Table 1 summarizes principal characteristics of the included studies in terms of demographics, the applied psychotherapy and a brief summary of the principal results as reported in each individual study. Furthermore, two authors have independently assessed the quality of the included studies applying the above-described NOS (New Castle–Ottawa Scale). All the studies included 1,688 participants with a range comprised between 18 and 65, and a median of 34.6 years old. Most of the studies ($n = 28$) applied cognitive-behavioral therapy (CBT) or non-psychodynamic, whereas specific PDT has been used in 4 studies, and specific mindfulness-based therapy in 6 sources. To assess the brain regions involved after a specific psychotherapy training, we conducted 3 different ALE meta-analyses (Table 2). The map including all the studies (ALL map) that assessed the longitudinal differences in brain activity, showed two homogeneous clusters in correspondence of left inferior frontal gyrus, concerning the orbitofrontal gyrus, and caudally involving the anterior portion of the insular cortex ($p < 0.0001$, corr.). Similarly, studies that assessed the longitudinal changes psychotherapy-related using emotional or cognitive tasks (TASK map) showed a left sided homogeneity in correspondence of the anterior insula ($p < 0.000$) extending to the Broca's area of the inferior frontal gyrus ($p < 0.0001$) and the superior frontal gyrus ($p < 0.0001$). Significant corrected results were not observed for the studies using resting state. After carrying

TABLE 1 Characteristics of the included studies.

References	Nr. Subjects (nr. of men)	Age (Mean years $\pm$ S.D.)	Diagnostic criteria	Rest vs. task	Therapy	MRI field	fMRI results
Buchheim et al. (2012)	MDD: 16 (3); HC: 17 (4)	38.9 $\pm$ 12.4	DSM-IV	Task: Adult Attachment Projective Picture System used to activate AA system	PDT	3T	Interaction between Group $\times$ Time resulted in activation of left amygdala, hippocampus, STG, vACC, MPFC
Wiswede et al. (2014)	MDD: 18 (4); HC: 17 (3)	MDD: 39.8 $\pm$ 12.8 HC: 38.0 $\pm$ 11.6	DSM-IV	Task: Operationalized Psychodynamic Diagnostics (OPD), traffic and relaxation sentences	PDT	3T	At T2 group $\times$ condition interaction, activation was observed in MFCG/ACC. No activation for OPD in MDD at T2.
Beutel et al. (2010)	PD: 9 (3); HC: 18 (9) 95	PD: 32 HC: 29	DSM-IV; ICD-10	Task: Emotional linguistic go/nogo task	PDT	1.5T	PD showed activation in SMA and lateral PFC for both emotional contexts. For positive Nogo > Positive go, at S2, PD showed increased activation in caudate nucleus.
Vuper et al. (2021)	PTSD: 26 ITT: 42 TEC: 18	PTSD: 33.62 $\pm$ 11.09 ITT: 31.17 $\pm$ 9.71 TEC: 31.89 $\pm$ 9.84	DSM-5	Rest ICA	CBT	3T	At T2 PTSD compared to TEC decreased connectivity between DMN and temporal and occipital regions. Changes in positive affectivity was related to connectivity between DMN, MTG and Lateral occipital cortex.
de Greck et al. (2013)	SD:15 (7) HC: 15 (7)	SD: 42.6 HC: 37.0	DSM-IV	Task: 1) Reward anticipation task; 2) empathy task.	PDT	1.5T	Bilateral hippocampus and left ITG showed augmented hemodynamic response. ROI-based results: For anger vs. control condition significant increase in post-central gyrus, STG, bilateral hippocampus, amygdala and insula
Dichter et al. (2009)	MDD: 12 (6) HC: 15 (6) 238	MDD: 39.0 $\pm$ 10.4 HC: 30.8 $\pm$ 9.6	DSM-IV	Task: The Wheel of Fortune (WOF)-decision-making task with selection, anticipation and feedback phases.	BATD	4T	Selection phase: Group $\times$ Time interaction showed increased activation in the MDD at T2 the paracingulate gyrus, the left putamen, the right supramarginal gyrus, and the left posterior temporal. Anticipation phase: MDD at T2: the left caudate, ACC, the left MFG and SFG, the left lingual gyrus, the left lateral and superior-lateral occipital cortex, the left Posterior parahippocampal gyrus, the right insula, right precuneus, right subcallosal cortex, right posterior temporal fusiform cortex, and bilateral precentral gyrus and temporal poles. Feedback phase: MDD at T2 were in the left planum temporale, right superior lateral occipital cortex, and right posterior temporal fusiform cortex.

(Continued)

TABLE 1 (Continued)

References	Nr. Subjects (nr. of men)	Age (Mean years $\pm$ S.D.)	Diagnostic criteria	Rest vs. task	Therapy	MRI field	fMRI results
Dichter et al. (2010)	DD: 12 (6) HC: 15 (6)	MDD: 39.0 $\pm$ 10.4SD HC: 30.8 $\pm$ 9.6	DSM-IV	Task: Sad vs. neutral images forced 2-choice task.	BATD	4T	Group $\times$ Time vs. Contrast Target Sad > Neutral showed increased activation at T2 in MDD in paracingulate gyrus
Fu et al. (2008)	MDD: 16 (3) HC: 16 (3)	MDD: 40.0 $\pm$ 9.4 HC: 39.20 $\pm$ 9.3	DSM-IV	Task: Sad faces morphed to represent 3 intensities. Subject had to indicate the gender of the faces	CBT	1.5T	ACC right MFG, right insula/IFG, and putamen showed a negative correlation with clinical response at T2. MDD showed a significant increase at T2 in : ACC, SFG, PCC/precuneus, inferior Parietal cortex.
Ritchey et al. (2011)	MDD: 11 (3) HC: 7 (2)	MDD: 36.1 $\pm$ 10.1SD HC: 34.6 $\pm$ 6.9	DSM-IV	Task: Emotion evaluation task (negative, neutral and positive stimuli) pleasantness assessment on a 3-point scale	CBT	1.5T	T1 differences between MDD and HC: reduction in vmPFC activity in MDD for the valence of the stimuli. T2: MDD increased activity of vmPFC.
Sankar et al. (2018)	MDD: 16 (3) HC: 16 (3)	MDD: 40 $\pm$ 9.27SD HC: 39.94 $\pm$ 9.48	DSM-IV	Task: mDAS-48 task (Dysfunctional Attitude Scale)	CBT	1.5T	MDD and HC showed a decrease in activation in the left parahippocampal gyrus at T2. In MDD, a main effect of time was observed in the right PCC. MDD: positive correlation between HAMD score and left precentral gyrus
Fonzo et al. (2017)	PTSD: 36 (13) HC: 30 (10) 413	PTSD: 34.42 $\pm$ 10.23SD HC: 39.03 $\pm$ 10.35	DSM-IV	Task : Emotion reactivity task; Emotional Conflict Task; Reappraisal Task	CBT; PE	3T	Reappraisal task: Increased activation in PTSD treatment group in MFG for condition decrease negative vs. look negative.
Fonzo et al. (2021)	PTSD: 36 (13) HC: 30 (10)	PTSD: 34.42 $\pm$ 10.23SD HC: 39.03 $\pm$ 10.35	DSM-IV	Rest Seed based connectivity	CBT; PE	3T	Significant treatment related increase in the amygdala connectivity with ventral anterior insula, DLPFC, vmPFC. Increased insula-amygdala connectivity at T2 of PE.
Huang et al. (2014)	MDD: 23 (7) HC: 20 (8)	MDD: 27.7 $\pm$ 10.9; HC: 28.8 $\pm$ 8.7	DSM-IV	Rest ReHo	CT	3T	At T1 MDD showed lower ReHo in the MPFC and rACC than HC. At T2 MDD showed no difference with HC.
Siegle et al. (2012)	MDD: 46 (9) HC: 35 (12) 4 cohort	Cohort 1 = 37.41 $\pm$ 9.17SD. Cohort 2 = 36.13 $\pm$ 11.11SD. Cohort 3 = 38.93 $\pm$ 9.74SD. Cohort 4 = 32.95 $\pm$ 9.37SD	DSM-IV	Task: Personal relevance rating task (PRRT).	CT	3T	Subgenual ACC activity was strongly related to changes and a more negative BDI in cohort 1.

(Continued)

TABLE 1 (Continued)

References	Nr. Subjects (nr. of men)	Age (Mean years $\pm$ S.D.)	Diagnostic criteria	Rest vs. task	Therapy	MRI field	fMRI results
Yoshimura et al. (2014)	MDD: 23 (16) HC: 15 (8)	37.3 $\pm$ 7.2	DSM-IV	Task: judgment tasks with self-reference, other-reference, semantic processing and letter-processing conditions.	CBT	1.5T	At T2 for group $\times$ time $\times$ valence, MDD showed increased activation for self/positive condition in vACC, STG, MPFC. Similar results for self/negative condition. Increased activation for MDD between T1 and T2. Positive correlation between rumination scores and vACC at T2 for self/negative
Nakao et al. (2005)	OCD: 10 (4) HC: 13 (6)	OCD: 32.4 $\pm$ 9.9SD; HC: 30 $\pm$ 6.4	DSM-III-R	Task: 1) Stroop task; 2) Symptom provocation task	BT; fluvoxamine	1.5T	Stroop task: At T2 OCD showed increased activation in bilateral prefrontal cortices, bilateral ACC, parietal cortices, and cerebellum. Symptom provocation task: At T2 OCD showed activation in OFC contrasting to Stroop task.
Bor et al. (2011)	SZ: 8 (6) HC: 15 (10) 705	SZ: 30.5 $\pm$ 8.3; HC: 28.5 $\pm$ 7.2	DSM-IV-TR	Task: N-back, 0-back and 2-back	CRT	1.5T	At T2, SZ during the spatial WM showed activation within the left vIPFC, the left ACC and left parietal areas.
Haut et al. (2010)	SZ: 18 (13) HC: 9 (7) 3 groups (REM, CBSST and Controls)	REM: 36.4 $\pm$ 9.2; CBSST: 39.5 $\pm$ 7.7; HC: 40 $\pm$ 6	DSM-IV	Task: 1) Word “N-back” 2) Picture “N-back” 3) Lexical decision Task.	REM; CBSST	3T	N-Back (word and picture): REM showed a number of regions with a significant activation > CBSST and HC: ACC, frontopolar, DLPFC. Lexical decision: no sig. results
Kumari et al. (2011)	SZ: 56 CBT + TAU group: 28 (18); TAU group: 28 (14)	SZ-CBT+TAU: 35.68 $\pm$ 7.82; SZ-TAU: 39.19 $\pm$ 9.37	DSM-IV	Task: Emotion recognition (sad, angry, happy and neutral faces)	CBT; Pharmacological	1.5T	Group $\times$ Time effect was found in bilateral IFG, right insula, bilateral putamen, left thalamus and left occipital areas (Fearful faces). CBT+ TAU showed decreased activation at aT2
Penadés et al. (2013)	SZ: 35 (27); HC: 15 (10)	SZ-CRT: 36.35 $\pm$ 13.6SD SZ- SST: 37.56 $\pm$ 8.99 HC: 34.75 $\pm$ 3.14	DSM-IV	Task: N-back task (0-back and 2-back)	CRT or SST	3T	Threatened patients showed a reduction in overactivation of the CEN during task-related responses. Patients showed a reduction in deactivation of DMN. CRT group: increase in white matter in the genu of corpus callosum.
Nabeyama et al. (2008)	OCD: 11 (4) HC: 19 (8)	OCD: 32.4 $\pm$ 10.6; HC: 32.7 $\pm$ 7.1	DSM-III-R	Task: Stroop task.	E/RP	1.5T	At T2 OCD exhibited significantly decreased activation in the right OFC, left MFG, left fusiform gyrus, bilateral parahippocampal gyrus, and left parietal lobe; increased activation in the right parietal lobe

(Continued)

TABLE 1 (Continued)

References	Nr. Subjects (nr. of men)	Age (Mean years $\pm$ S.D.)	Diagnostic criteria	Rest vs. task	Therapy	MRI field	fMRI results
Goldin and Gross (2010)	SAD: 31 (NA) HC: 29(NA)	NA	DSM-IV	Task: Personally-salient autobiographical social situations	CBT	3T	Group $\times$ time allowed involvement of DLPFC and DMPFC in CBT group.
Ma et al. (2021)	OCD: 59 (22)	OCD-CCT: 27.2 ± 5.7 OCD-CCT-SSRI: 28.1 ± 5.1 OCD-SSRI: 28.1 ± 5.1	DSM-IV	Rest ALFF	CCT; pCCT/p	3T	No significant clusters (ALFF) at T2 in the 3 groups of OCD patients.
Zhu et al. (2018)	PTSD: 25 (7) HC: 26 (7)	PTSD: 35.3 ± 10.4 SD; HC: 35.9 ± 8.9 SD	DSM-IV-TR	Rest Seed based connectivity	PE	1.5T	Interaction group $\times$ time showed significant FC of the basolateral amygdala with OFC, vmPFC and thalamus.
Yoshino et al. (2018)	SD: 29 (9) HC: 31 (5)	SD: mean 46.7 ± 11.3 ; HC: 46.9 ± 10.3	DSM-IV	Rest ICA	Group CBT	3T	After CBT Intrinsic connectivity network (ICN) increased in OFC of DAN; decreased in right IPL of DAN and left PCL of SMN. At T2 negative correlation between VAS and right OFC. Vup
Sheynin et al. (2020)	PTSD: 61 (61) HC: 29 (13)	PT: $= 35.23 \pm$ S.D. $= 8.6$ HC $= 32.38$ S.D. $= 8.27$	DSM-IV- TR	Rest Seed based connectivity ROI-ROI connectivity	PE	3T	Statistical trend in the Amygdala-PCC connectivity for PTSD treatment responders.
Yuan et al. (2018)	SAD: 15 (10) HC: 19 (13)	SAD: 27.07 ± 8.11 SD; HC: 26.6 ± 4.9 SD	DSM-5	Rest ALFF	Group CBT	3T	Compared to T1, at T2 SAD showed decreased DC in left precuneus and MTG and increased in putamen. Positive correlation between DC in precuneus and LSAS.
Yuan et al. (2016)	SAD: 15 (10) HC: 19 (13)	SAD: 27.07 ± 8.11 SD; HC: 26.6 ± 4.9 SD	DSM-5	Rest Seed based connectivity	Group CBT	3T	Compared to T1, SAD abnormal hyperconnectivity between Amygdala and dACC, and improved clinical symptoms.
Pantazatos et al. (2020)	MDD:20 (12) HC: 30 (19)	MDD: 34.1 ± 10 ; HC: 32.4 ± 10	DSM-IV	Rest Seed based connectivity	CBT	3T	At T2, there was decreased connectivity in both rACC and BA25 and PCL, resulting in a more normalized pattern of activity.
Zhao et al. (2021)	OCD: 79 (32) 3 groups	Group 1: 30.8 ± 9.2 SD. Group 2: 30.2 ± 9.4 SD. Group 3: 29.1 ± 7.3 SD.	DSM-IV	Rest Seed based connectivity	CCT; pharmacotherapy	3T	OCD under CCT treatment showed decreased FC of left Amygdala with right ACC and left precentral lobe.
Leroy et al. (2022)	PTSD: 30 (15) 2 groups	PTSD-responders: Mean $= 41.3 \pm 12.3$ SD PTSD-non responders: Mean $= 36.6 \pm 13.5$ SD	DSM-IV-TR	Rest FC assessed with Granger causality analysis	TMRT	3T	Significant changes in Anterior Insula causal maps after TMRT.

(Continued)

TABLE 1 (Continued)

References	Nr. Subjects (nr. of men)	Age (Mean years $\pm$ S.D.)	Diagnostic criteria	Rest vs. task	Therapy	MRI field	fMRI results
Sculth et al. (2019)	GAD: 25 (9)	GAD: 21.8 $\pm$ 2.6	DSM-IV	Rest Seed based connectivity	ERT	3T	T2 > T1: 1) PCC seed: increased connectivity with middle occipital gyrus, precuneus cuneus, precentral gyrus/motor cortex and premotor areas/DLPFC. 2) Anterior insula seed: increased connectivity with precuneus. 3) Posterior insula seed: increased connectivity with anteromedial PFC/dACC and decreased connectivity with midbrain.
King et al. (2016)	14 PTSD: 14 (NA) PCGT: 9 (NA)	PTSD: 32.43 $\pm$ 7.54SD PCGT: 31.67 $\pm$ 10.14	DSM-IV	Rest: Seed based connectivity Task: Emotion regulation task	MBET	3T	At T2: Increased PCC connectivity with bilateral DLPFC and dACC and between left amygdala, left hippocampus and dACC.
Zhao et al. (2019)	GAD: 32 (24)	33.62 $\pm$ 7.71SD	DSM-IV	Rest ReHo	MBCT	3T	Decreased ReHo in the right middle temporal gyrus, bilateral insula, bilateral STP, right MCC and left hippocampus; Increased PCC connectivity with ACC.
Goldin et al. (2009)	SAD: 25 (16)	35.2 $\pm$ 11.9	DSM-IV	Task: SRP task	MBSR	3T	Positive SRP: decreased BOLD response in dorsomedial e medial PFC, left inferior frontal gyrus. Negative SRP: increased BOLD response
Cernasov et al. (2021)	Anhedonia Subject: 73 (21) 2 groups (BATA and MBCT)	MBCT: 31.8 $\pm$ 9.2SD BATA: 27.9 $\pm$ 8.8	DSM-5	Rest ROI-to-ROI network connectivity	BATA and MBCT	7T	No significant between group differences.
Bachmann et al. (2018)	ADHD: 40 (NA) 2 groups (MAP and PE)	MAP: 40 $\pm$ 10.58SD; PE: 40.26 $\pm$ 13.81	DSM-IV	Task: Working memory task	MAP or PE	3T	MAP: activations in bilateral inferior parietal lobule, posterior insula and precuneus.
Goldin and Gross (2010)	SAD: 16 (7)	35.2 $\pm$ 11.9	DSM-IV	Task: Emotion regulation task	MBSR	3T	Activations in inferior and superior parietal lobule, precuneus, cuneus, middle occipital gyrus and parahippocampal gyrus.

HC, Healthy control; Dx, Diagnosis; MDD, Major Depression Disorder; PD, Panic Disorder; PTSD, Post-Traumatic Stress Disorder; SZ, Schizophrenia; SAD, Social Anxiety Disorder; SD, Somatoform Disorder; GAD, General Anxiety Disorder; OCD, Obsessive Compulsive Disorder; Anhe, Anhedonia; PDT, Psychodynamic Therapy; CBT, Cognitive Behavioral Therapy; BATD, Behavioral Activation Therapy for Depression; EP, Exposure Therapy; WL, Waiting List (not healthy; for instance [Fonzo et al., 2021](#)); BT, Behavioral therapy; CRT, Cognitive remediation therapy; CBSST, cognitive behavioral social skills training; REM, remediation training; TAU, Treatment as usual; SST, social skills training; MT, Mindfulness training; MBSR, Mindfulness-based stress reduction; E/RP, exposure and response prevention; NSBs, reappraising self-generated negative self-beliefs; MBET, mindfulness-based exposure therapy; MBCT, Mindfulness-based cognitive therapy; MBM, Mindfulness-based meditation; HEP, Health enhancement program; SRP, Self Referential Processing; MAP, Mindfulness Awareness Practice; CCT, Coping Cognitive Therapy; pCCT, pharmacological CCT; p, Pharmacological therapy; PE, Prolonged Exposure Therapy; BATA, Behavioral Activation Therapy for Anhedonia; TMRT, traumatic memory reactivation therapy; ERT, Emotion regulation therapy; ICA, Independent Component Analysis; ReHo, Region Homogeneity; ROI, Region of Interest; FC, Functional Connectivity; ALFF, Analysis Low Frequency Fluctuation.

TABLE 2 Brain clusters resulting from ALE meta-analysis performed for all psychotherapy, CBT, PDT, and mindfulness based studies.

Cluster	BA	Hemisphere	<i>x</i>	<i>y</i>	<i>z</i>	ALE	<i>P</i>	<i>Z</i>
All psychotherapy studies								
Inferior frontal gyrus/anterior insula	13	L	−44	26	8	0.021	0.000009	4,285
Inferior frontal gyrus/ orbital gyrus	47	L	−42	22	2	0.020	0.000012	4,232
CBT								
Anterior insula	13	L	−44	26	8	0.01977	0.00001	4.41716
Anterior insula	13	R	38	18	4	0.01728	0.00003	4.04779
Superior frontal gyrus	8	L	0	28	48	0.01783	0.00002	4.13672
Psychodynamic psychotherapy								
Inferior frontal gyrus	47	R	21	21	−12	0.008569	0.000093	3.736320
Lentiform nucleus/putamen		L	−25	−12	7	0.008569	0.000093	3.736320
Superior frontal gyrus	8	R	3	44	49	0.008706	0.000053	3.875660
Anterior cingulate cortex	32	R	1	46	4	0.009100	0.000012	4.215530
Mindfulness								
Inferior frontal gyrus	47	L	−44	22	0	0.01458	0.00000	4.55147

BA, Broadman Area; L, Left; R, Right.

TABLE 3 Brain clusters resulting from the conjunction analysis between resting state and task based studies.

Cluster	BA	Hemisphere	<i>x</i>	<i>y</i>	<i>z</i>	<i>p</i>	<i>Z</i>
Conjunction analysis REST vs. TASK							
dACC	32	R	4	32	22	0.0054	2,549
ACC	24	L	−2	30	22	0.0056	2,536
ACC	24	L	−1	32	15	0.0102	2,319
ACC	24	L	−8	32	18	0.0126	2,238
dACC	32	L	−12	22	18	0.0142	2,192
ACC	24	L	−4	18	22	0.0182	2,092
ACC	24	L	−8	24	14	0.0192	2,071
ACC	24	L	−10	18	20	0.027	1,927
Medial frontal gyrus	11	R	9	28	−12	0.0478	1,667
Putamen	–	R	24	0	−4	0.0384	1,770
Middle temporal gyrus/angular gyrus	39	R	48	−68	18	0.0432	1,715
Middle temporal gyrus/angular gyrus	39	R	46	−68	20	0.0474	1,671
Middle temporal gyrus/angular gyrus	39	R	44	−66	20	0.0474	1,671

BA, Broadmann Area; L, Left; R, Right; dACC, Dorsal anterior cingulate cortex; ACC, anterior cingulate cortex.

out a conjunction analysis between resting state and task-based studies, for the resting state studies, uncorrected results ($p < 0.006$; $p < 0.05$) have been observed in correspondence of anterior cingulate cortex (ACC) and dorsal ACC ($p < 0.006$). Moreover, a small homogeneity cluster was found in the right Putamen (Table 3 and Figure 2). These results about resting state studies partially overlay the anatomofunctional localization of the one of the hubs of the default network (DN), located in vmPFC/ACC since the extracted coordinates were about FC results. Similarly, our interest was in studying the contribution of each type of psychotherapy to the brain functional response. Interestingly, the studies that used psychodynamic psychotherapy showed Family-Wise Error

(FWE) cluster-corrected ($p < 0.05$) homogeneity values in correspondence of the right superior and inferior frontal gyri, with a small cluster in the putamen (Table 3).

No FWE-corrected homogeneity foci have been observed for Mindful- based and CBT psychotherapy. Comparing T1 and T2, CBT studies showed two bilateral clusters in the dorsal anterior insulae ($p < 0.001$) and the involvement of the medial superior frontal gyrus ($p < 0.0005$). Moreover, after Mindful-based therapy, left inferior frontal gyrus showed uncorrected homogeneity ($p < 0.0001$).

Furthermore, Behavioral Analysis (Lancaster et al., 2012) of the ALE resulting maps has been performed using Behavioral Analysis plugin of Mango v.2.6 (<http://ric.uthscsa.edu/mango/>)

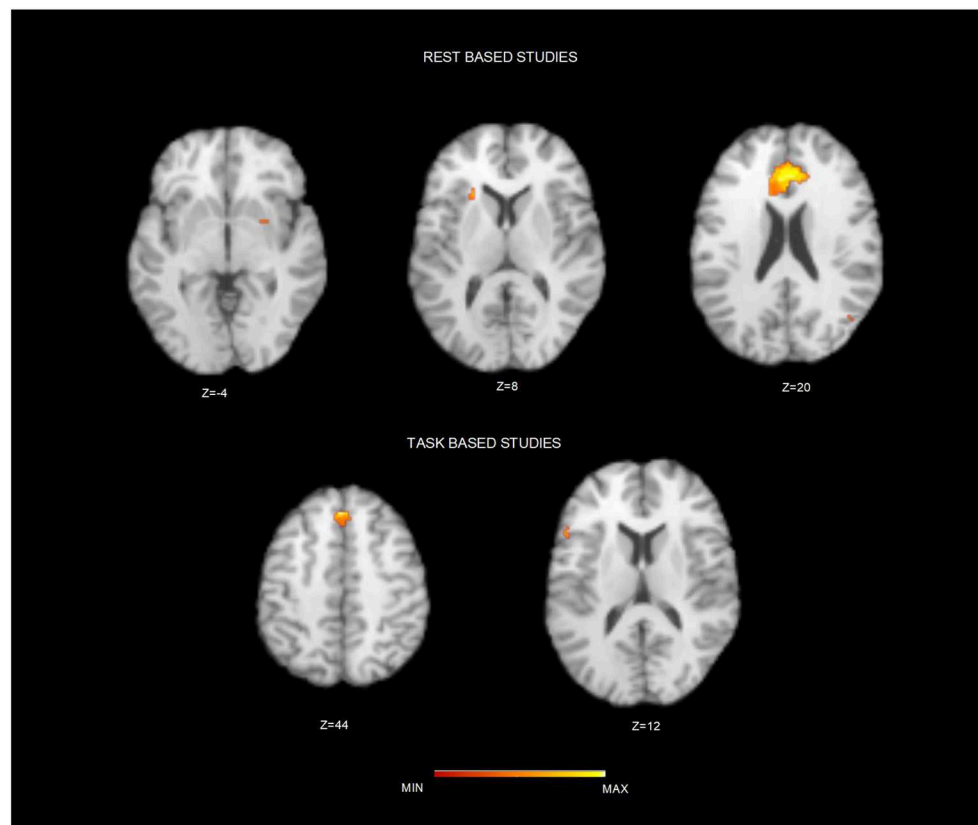


FIGURE 2

Brain clusters resulting from the conjunction analysis between resting state and task-based studies. Figure depicts the resulting brain cluster after applying the conjunction analysis between resting state (**Top**) and task-based studies (**Down**). Maps are over imposed on a Talairach template in the axial plane and in neurological convention.

[plugin_behavioralanalysis.html](#)). Behavioral Analysis presented for BrainMap's five Behavioral Domains (Action, Cognition, Emotion, Interoception, and Perception) and sixty sub-domains. Only z-scores ≥ 3.0 are considered significant ($p \leq 0.05$ with Bonferroni corr., Figure 3). All psychotherapies resulted in a map that allowed significant Behavioral analysis for the domain cognition (Language and Memory; $z = 5.09$) and emotion (Negative; $z = 3.04$). ALE task map did not show significant results for the above-mentioned domains. ALE resting state fMRI map showed significant results for the domain emotion (Positive reward /Gain; $z = 3.236$). The observed homogeneous clusters in the CBT-ALE map allowed significant cognition domain (reasoning, language, memory, and attention) and emotion domain (reward).

Discussion

In the present study we sought to delineate the neural effects of shared and unique effects of different psychological therapies, evaluating the findings of existing longitudinal prospective

studies in different psychopathological conditions. In other words, we tried to answer the question of how brain function changes after psychotherapy, using fMRI. Our second, related question was to assess the contribution of psychodynamic therapy from a neural point of view, underscoring any potential differences with non-PDT.

Our meta-analysis consisted of studies that had acquired serial neuroimaging scans prior to and following a course of treatment with a psychological therapy. Longitudinal brain imaging studies of both resting state and of emotional-cognitive tasks, using fMRI, were included, and analyzed separately and conjunctly.

All studies together

The map including all studies that assessed longitudinal brain changes showed two homogeneous clusters from the orbitofrontal cortex (OFC), and caudally involving the anterior portion of the insular cortex. This finding is not surprising, both because of previous similar results existing in the literature

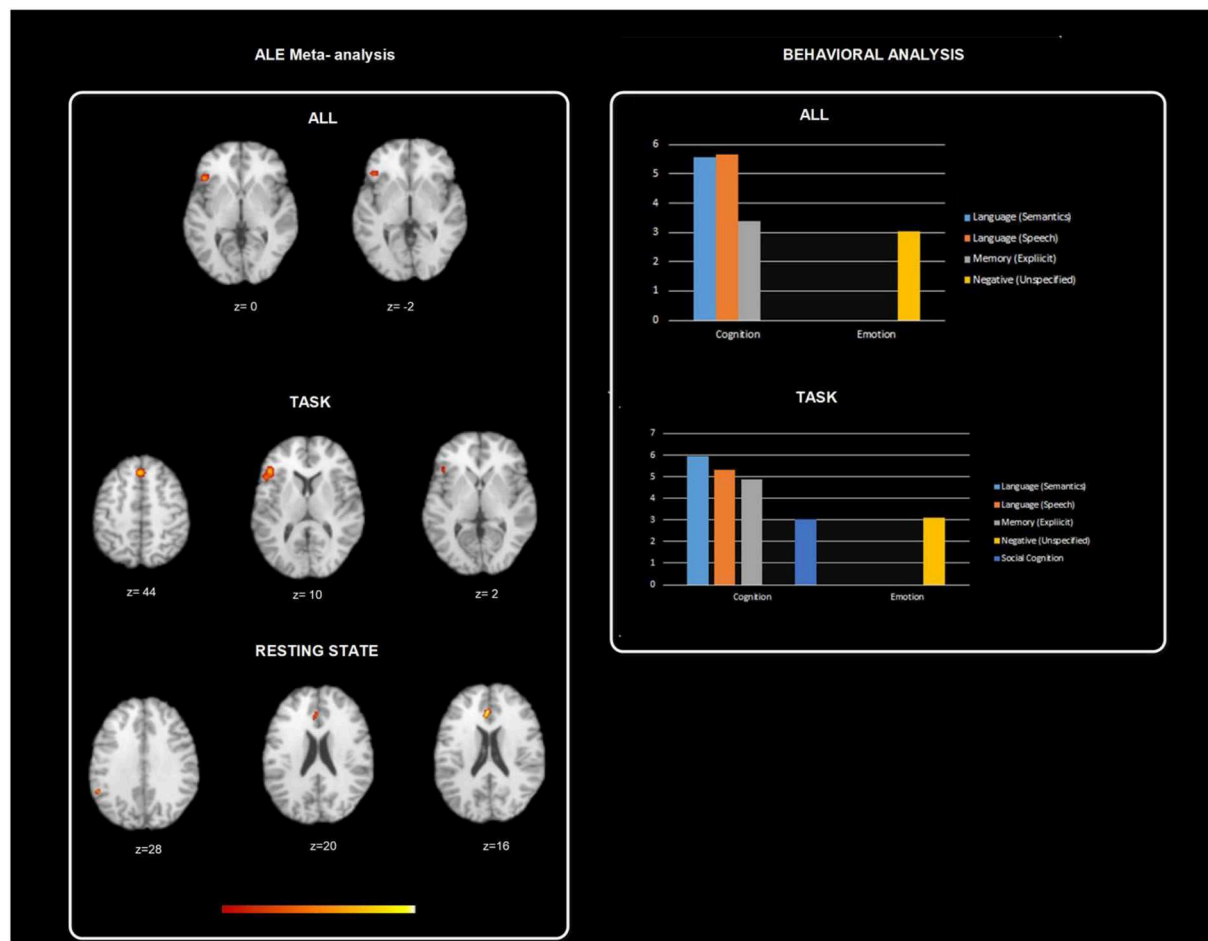


FIGURE 3

Figure depicts behavioral analysis of the ALE resulting maps using Behavioral Analysis plugin of Mango v.2.6. Behavioral Analysis presented for BrainMap's five Behavioral Domains (Action, Cognition, Emotion, Interoception, and Perception) and sixty sub-domains. Only z-scores ≥ 3.0 are considered significant ($p \leq 0.05$ with Bonferroni corr.). All psychotherapies resulted in a map that allowed significant behavioral analysis for the domain cognition (Language and Memory; $z = 5.09$) and emotion (Negative; $z = 3.04$). ALE task map did not show significant results for the above-mentioned domains. ALE resting state fMRI map showed significant results for the domain emotion (Positive reward /Gain; $z = 3.236$). The observed homogeneous clusters in the CBT-ALE map allowed significant cognition domain (reasoning, language, memory, and attention) and emotion domain (reward).

and the specific functions of these regions. Specifically, Fu et al. (2013) found the OFC is one of the regions predictive of a good clinical response to psychotherapy. The prefrontal cortex, in general, has a complex role in the control and organization of behavior. In this context, the OFC is a fundamental part of the mesolimbic system, uniquely placed to integrate sensory and autonomic information to modulate behavior through both visceral and motor systems (Kringelbach, 2005). This structure is specialized in processing the reward value of numerous types of stimuli (Rolls et al., 2020). Greater activity in this area may reflect an increased responsivity to hedonic stimuli and reward, the inverse of anhedonia, which is predictive of a better clinical response (Fu et al., 2013). The OFC also has a role to connect and “make sense” of reality through sensory integration, modulation

of visceral reactions, participating in learning, prediction and decision making for emotional and reward-related behaviors (Kringelbach, 2005). All these complex activities are critical for psychotherapy, from the embodied mind to emotion, learning, and prediction. Decreased functional connectivity of the OFC with medial temporal lobe areas involved in memory is found in patients with depression. Rolls et al. (2020) point out some therapies for depression could increase the activity or connectivity of the medial OFC.

The result at the insular cortex is even less surprising since this region conveys sensory information, in contact with the external reality, to further brain regions that allow processing. Insula is anatomically situated in a brain area connected with several functional neural circuits supporting

cognitive, homeostatic, and affective systems. Its position and function represent a bridge between brain regions involved in monitoring internal states (visceral sensory, somatic sensory processes, autonomic regulation of the gastrointestinal tract and heart; Singer et al., 2004; Menon and Uddin, 2010; Esposito et al., 2018), supporting their processing. This pathway is called the “homeostatic afferent pathway” (Craig, 2009) carrying information about the body. Particularly, information arising from the body reaches the middle and posterior parts of the insula and then is projected in the anterior insula. The awareness of salient events is represented in the anterior insula, whereas more sensory attributes are represented posteriorly (Craig, 2002, 2003, 2009). Insula is a core structure that receives bodily information, filters salient stimuli, processes them, and then engages—through anterior cingulate cortex (ACC), the central executive network, and the default network (DN)—memory and functions connected to the self (Cieri and Esposito, 2019; Cieri, 2022). Given its functions, the insula is considered a central hub for the allostatic-homeostatic regulatory process between mother and infant (Scalabrini et al., 2022). This element makes considerable sense in a psychodynamic context, where the transferal phenomena between the analyst and the analysand can recall the mother and the infant. This region is also considered a bridge that connect the three level of self as recently proposed by Northoff and colleagues (Scalabrini et al., 2018, 2022; Qin et al., 2020).

Emotional and cognitive tasks

The result at the anterior insula is also present in emotional and cognitive task studies, together with the inferior and superior frontal gyrus. Within the frontal lobe, inferior and superior frontal gyrus are the target influenced especially by PDT and mindfulness, while CBT was linked to significant modification of medial superior frontal gyrus. The inferior frontal gyrus corresponds to Broca's area, therefore involved in language processing, speech production, lexical and semantic processes, syntactic, and phonological processes, all functions present and vital in the context of the talking cure. Moreover, other fMRI research has shown that the blood oxygenation level dependent (BOLD) signal within the inferior frontal gyrus increases at the point of inhibitory control when compared to a baseline of routine responding (Menon et al., 2001). In other words, this area seems to play a key role in the inhibition processes (Verbruggen and Logan, 2008). This result makes sense in the task paradigm, and it also finds a natural expression within a psychotherapeutic treatment where the inhibition processes are often faced and discussed within treatment.

The superior frontal gyrus is thought to contribute to higher cognitive functions and particularly to working memory, but its activation during conflict anticipation is positively

correlated with the capacity of inhibitory control associated with both efficient response inhibition and less motor urgency (Hu et al., 2016). Again, we find an involvement of inhibitory control through top-down processes or more focus on bottom-up mental states. We will take up these results about the superior and inferior frontal gyrus later in the discussion of the PDT.

Resting state studies

Although significant corrected results were not observed for the studies using resting state, we will briefly discuss the uncorrected results at the level of another important area: the ACC. As we pointed out in a prior study (Cera et al., 2019) this structure is involved along with the insular cortex, the secondary somatosensory cortex, the nuclei in the tegmentum and the hypothalamus, in the regulation of attentional focus by integrating external and internal stimuli, and in the expression of emotional states, thus modulating a motivational state toward homeostasis (Damasio et al., 2000; Cera et al., 2019). Sankar et al. (2018), in their meta-analysis showed significant interaction effect of CBT to the rostral ACC in depressed patients compared to healthy controls. The findings could provide some insight into the potential mechanisms and specificity of treatment effects of this therapy. Patients with major depression showed increased activity following psychological therapy while healthy subjects have shown decreased activity at the follow up scan. In our current result, we found the involvement of the dorsal ACC, part of the salience network, implicated with emotional processing and the supramarginal gyrus, part of the DN, which plays a role in the therapeutic process (Cieri, 2022; Rabeyron, 2022). The poor specificity of our study (disorders and therapies) and the relatively small sample size could be among the reasons why the ACC did not survive multiple comparisons correction.

Psychodynamic specificity

We also were interested in addressing the specific contribution of PDT. While the low number of psychodynamic studies in this field limits statistical power, the studies that used PDT showed Family-Wise Error (FWE) cluster-corrected ($p < 0.05$) homogeneity values in the right superior and inferior frontal gyri, with a small cluster in the putamen. We have already mentioned the potential role of the inferior and superior frontal gyri, extending beyond the important function of language to the potential inhibitory role stimulated or developed in therapy. We also note that the inferior frontal gyrus, as part of the frontal cortex, is a neocortical region that coordinate a wide range of neuropsychological processes (Miller and Cohen, 2001), with an important role modulating bottom-up process such as regulation of behavior that is more automatic, but also

included in processes when behavior must be guided by internal states, with a more classic top-down function. Many regions in this area overlap with the human mirror neuron network (inferior frontal gyrus and superior temporal gyrus) involved in action observation and execution (Rizzolatti and Craighero, 2004). Modulation, inhibition processes, focus on internal states and the involvement of the mirror neuron network are vital components of any psychotherapy, in fact in this case a result present in PDT, CBT and mindfulness. The insula also has been shown to act in concert with the human mirror neuron system during imitation and observation of emotions (e.g., empathy; Carr et al., 2003; Uddin et al., 2009), which again, are expected to have a fundamental role in psychotherapy, not only as a requirement of the therapist and the patient within the therapy, but as a function potentially developed during treatment.

Concerning the insula, it is worth mentioning that the anterior portion of this region (and the von Economo neurons it contains) has a crucial role in awareness, and thus it needs to be considered as a potential neural correlate of consciousness (Craig, 2009) with a peculiar meaning for our study, and with a potential fundamental role in psychotherapy.

A specific region significant for the psychodynamic approach is the putamen. It is part of basal ganglia, which has long been assumed not only to play a role in motor planning and control, but also involved in several language aspects (Viñas-Guasch and Wu, 2017) including lexical, morphological (Friederici, 2002), syntactic (Teichmann et al., 2015), and speech production processes (Oberhuber et al., 2013). Moreover, the basal ganglia are involved in mammalian learning and memory (Aosaki et al., 1994; for a review see Packard and Knowlton, 2002). It should be noted that there is also extensive research examining the role of the basal ganglia in adaptation in motor control (Graybiel et al., 1994), neural representations of habits (Jog et al., 1999), space and direction and navigation (Wiener, 1993; Mizumori et al., 2000), explored also through neural computational modeling (e.g., Gillies and Arbuthnott, 2000). We can speculate that the involvement of this region in PDT—with a role not only on movement and language, but also involved in habits, neural representation, navigation, and memory—could be stimulated by a technique that works more with space-time, memory and dreaming, compared to more cognitive psychotherapy approaches that are focused on symptoms and the here and now, with less focus to the past and future (Cieri, 2022), landmark of psychoanalysis and PDT. We also found it intriguing that some authors (Bartels and Zeki, 2004; Zeki and Romaya, 2008) found equal activation of the network involving this region by love and hate (Zeki and Romaya, 2008). We can speculate that this overlapping has several theoretical and clinical psychodynamic precedents. In “Instinct and Their Vicissitudes”, Freud (1915) claims that love and hate characterize the natural relationships of the ego with the objects. Klein (1937) at the very beginning of her “Love,

Guilt, and Reparation” points out how destructive impulses play in interaction of hate and love and how feelings of love and tendencies to reparation develop in connection with aggressive impulses and in spite of them. In fact, the baby’s first object of love and hate—her mother—is both desired and hated with all the intensity and strength that is characteristic of the early urges of the baby. From a more clinical perspective, Winnicott (1949) in “Hate in the Counter-Transference”, resume Freud and Klein’s perspectives, underscoring that these two affects overlap both in the patient, and the therapist. He points out that during the analysis of patients with and without psychosis, therapists must find themselves in a position comparable to that of the mother of a new-born baby, highlighting the presence of love and hate, both in the mother toward the baby and in the therapist toward the patient.

Before the conclusion and specifically with regard to the psychodynamic contribution, we want to use a metaphor from Kant (1781, 2003), from “Critique of Pure Reason” where he uses the metaphor of the dove to express something potentially similar to the dialogue between neuroscience and psychoanalysis:

“The light dove, in free flight cutting through the air the resistance of which it feels, could get the idea that it could do even better in airless space. Likewise, Plato abandoned the world of the senses because it posed so many hindrances for the understanding and dared to go beyond it on the wings of the ideas, in the empty space of pure understanding.”

This image of dove and its flight could be taken in consideration in this field thinking about some psychoanalysis that believes that it doesn’t need methodological research, especially in the field of the brain sciences. On the other hand, neuroscientific research nowadays relies almost exclusively on algorithms and artificial intelligence without taking into consideration the subjective experience of patients. Today, a psychodynamic approach without dialogue with neuroscience sounds “brainless.” In contrast, brain neuroscientific research sounds “mindless” without the consideration of subjective individual experience (Cieri, 2022). Each approach needs the other side of the mind-brain system; according to Kant’s metaphor, both sides need the resistance of the air to fly.

Conclusions

In this work we systematically reviewed the evidence of effect of psychological therapy on brain function, trying to better understand the neurobiological bases of the effectiveness of psychotherapeutic treatments. All the psychological approaches seem to influence the brain from a functional point of view, showing their efficacy from a neurological perspective. It is not easy to delineate a precise and distinguished pattern of

changes in specific disorders or theoretical approaches. Frontal, prefrontal regions, insular cortex, superior and inferior frontal gyri, and putamen seem involved in these neural changes, with the PDT more involved in the latest three regions.

Limitations

Our study has limitations, such as the subjectivity of the individuals as response to the therapy. This element, of course is not a limitation *per se*, it is impossible to capture in its complexity, and it can be applied to the efficacy of the therapy in general. As Fonagy (2015) considered, it might be understandable to wish for an intellectual short-cut to a pooled effect size rather than considering individual investigations, but meta-analyses lack individual patient data—they are based on response rates and mean values. Being aware of this aspect, in our study we tried to discuss neural changes due to a relationship between two individuals (as a psychotherapeutic setting). Even before the limitation derived from putting together different therapeutic approaches, there is an intrinsic and ineliminable limitation in the objectification of a relationship unique *par excellence*. Related to this limitation, the pre-post therapy observed brain changes could be related to some other factors, not necessary to the specific treatment received. As underlined by other colleagues (Messina et al., 2013; Franklin et al., 2016) such studies are characterized by heterogeneity of therapeutic approaches and study designs. We have differences in sample characteristics, such as the specificity of the disorder (e.g., proportion experiencing first depressive episode vs. recurrent illness), number and duration of sessions, regions chosen for reduced threshold analyses, scanner resolution and nature/existence of comparator group etc. that can be seen as potential confounders and limitations. As mentioned, some of these limitations are intrinsic and unavoidable. Another important limitation is the great variety of statistical analysis techniques used in previous resting state fMRI studies that we included. Since the methods to assess low frequency fluctuation, and their variation related to psychotherapy, of BOLD signal can be considered a bias in the study results. Indeed, we included several resting state studies that applied a seed-based connectivity analysis but using different seeds (Wolters et al., 2019). In this way, more studies applying the same analytical methods should be needed, allowing a more homogeneous and unbiased results as reported by previous meta-analysis (Iwabuchi et al., 2015). Finally, our sample size of the total studies used, is relatively small, particularly derived from the PDT little contribution. Specifically, as recalled by Eickhoff et al. (2016), cluster-level thresholding does a very good job of controlling excessive contribution of one experiment if 17 or more experiments are included in an ALE analysis.

In our case, we have used 18 studies, but only 4 studies from PDT.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

NC and JM: investigation, methodology, formal analysis, visualization, and writing review and editing. RE, GD, DC, and JC: visualization and writing review and editing. FC: conceptualization, investigation, methodology, visualization, writing original draft, and writing review and editing. All authors contributed to the article and approved the submitted version.

Funding

NC was supported by a Grant from FCT (FCT-PTDC/PSI-GER/30520/2017 and NORTE-01-0145-FEDER-030520). DC was supported by a Grant from the National Institute on Aging (RF1AG071566), an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Grant Number 5P20GM109025, and the Nevada Exploratory Alzheimer's Disease Research Center (NVeADRC; P20-AG068053). In addition, research reported in this publication was supported in part by private Grants from the Peter and Angela Dal Pezzo funds, from Lynn and William Weidner, and from Stacie and Chuck Matthewson. JC was supported by the Women's Alzheimer's Movement at Cleveland Clinic, National Institute on Aging Grants R01-AG074392 and P20-AG068053, and the COBRE funding: P20 GM109025. FC was supported by the Women's Alzheimer's Movement at Cleveland Clinic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or

claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abbass, A., Town, J., and Driessen, E. (2012). Intensive short-term dynamic psychotherapy: a systematic review and meta-analysis of outcome research. *Harv. Rev. Psychiatry*. 20, 97–108. doi: 10.3109/10673229.2012.677347
- Almohammed, O. A., Alsalem, A. A., Almangour, A. A., Alotaibi, L. H., Al Yami, M. S., Lai, L., et al. (2022). Antidepressants and health-related quality of life (HRQoL) for patients with depression: analysis of the medical expenditure panel survey from the United States. *PLoS ONE* 17, e0265928. doi: 10.1371/journal.pone.0265928
- America's State of Mind Report (2020). Available online at: <https://www.express-scripts.com/corporate/americas-state-of-mind-report> (accessed August 20, 2020).
- Aosaki, T., Graybiel, A. M., and Kimura, M. (1994). Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science* 265, 412–415. doi: 10.1126/science.8023166
- Bachmann, K., Lam, A. P., Sörös, P., Kanat, M., Hoxhaj, E., Matthies, S., et al. (2018). Effects of mindfulness and psychoeducation on working memory in adult ADHD: A randomized, placebo-controlled fMRI study. *Behav. Res. Ther.* 106, 47–56. doi: 10.1016/j.brat.05.002
- Barber, J. P., Barrett, M. S., Gallop, R., Rynn, M. A., and Rickels, K. (2012). Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *J. Clin. Psychiatry* 73, 66–73. doi: 10.4088/JCP.11m06831
- Barber, J. P., Muran, J. C., McCarthy, K. S., Keefe, J. R., and Zilcha-Mano, S. (2021). "Research on dynamic therapies," in *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change: 50th Anniversary Edition*, eds M. Barkham, W. Lutz, and L. G. Castonguay (John Wiley & Sons, Inc.), 387–419.
- Bartels, A., and Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage*. 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003
- Beutel, M. E., Stark, R., Pan, H., Silbersweig, D., and Dietrich, S. (2010). Changes of brain activation pre- post short-term psychodynamic inpatient psychotherapy: an fMRI study of panic disorder patients. *Psychiatry Res.* 184, 96–104. doi: 10.1016/j.psychres.06.005
- Bloch, M., Meiboom, H., Lorberblatt, M., Bluvstein, I., Aharonov, I., and Schreiber, S. (2012). The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of post- partum depression: a randomized, double-blind, placebo-control- led study. *J. Clin. Psychiatry* 73, 235–241. doi: 10.4088/JCP.11m07117
- Boeker, H., Richter, A., Himmighoffen, H., Ernst, J., Bohleber, L., Hofmann, E., et al. (2013). Essentials of psychoanalytic process and change: how can we investigate the neural effects of psychodynamic psychotherapy in individualized neuro-imaging? *Front. Hum. Neurosci.* 7, 355. doi: 10.3389/fnhum.2013.00355
- Bor, J., Brunelin, J., d'Amato, T., Costes, N., Suaud-Chagny, M. F., Saoud, M., et al. (2011). How can cognitive remediation therapy modulate brain activations in schizophrenia? An fMRI study. *Psychiatry Res.* 192, 160–166. doi: 10.1016/j.psychres.12.004
- Buchheim, A., Viviani, R., Kessler, H., Kächele, H., Cierpka, M., Roth, G., et al. (2012). Changes in prefrontal-limbic function in major depression after 15 months of long-term psychotherapy. *PLoS ONE* 7, e33745. doi: 10.1371/journal.pone.0033745
- Carhart-Harris, R. L., and Friston, K. J. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain*. 133, 1265–1283. doi: 10.1093/brain/awq010
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. USA* 100, 5497–5502. doi: 10.1073/pnas.0935845100
- Cera, N., Esposito, R., Cieri, F., and Tartaro, A. (2019). Altered cingulate cortex functional connectivity in normal aging and mild cognitive impairment. *Front. Neurosci.* 13, 857. doi: 10.3389/fnins.2019.00857
- Cera, N., Vargas-Cáceres, S., Oliveira, C., Monteiro, J., Branco, D., Pignatelli, D., et al. (2021). How relevant is the systemic oxytocin concentration for human sexual behavior? A systematic review. *Sex Med.* 9, 100370. doi: 10.1016/j.esxm.2021.100370
- Cernasov, P., Walsh, E. C., Kinard, J. L., Kelley, L., Phillips, R., Pisoni, A., et al. (2021). Multilevel growth curve analyses of behavioral activation for anhedonia (BATA) and mindfulness-based cognitive therapy effects on anhedonia and resting-state functional connectivity: Interim results of a randomized trial?. *J. Affect. Disord.* 292, 161–171. doi: 10.1016/j.jad.05.054
- Cieri, F. (2022). Memory for the future: psychodynamic approach to time and self through the default network. *Front. Hum. Neurosci.* 16, 885315. doi: 10.3389/fnhum.2022.885315
- Cieri, F., and Esposito, R. (2019). Psychoanalysis and neuroscience: the bridge between mind and brain. *Front. Psychol.* 10, 1790. doi: 10.3389/fpsyg.2019.01983
- Cieri, F., Zhuang, X., Caldwell, J. Z. K., and Cordes, D. (2021). Brain entropy during aging through a free energy principle approach. *Front. Hum. Neurosci.* 15, 647513. doi: 10.3389/fnhum.2021.647513
- Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., et al. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 391, 1357–1366. doi: 10.1016/S0140-6736(17)32802-7
- COVID-19 Mental Disorders Collaborators (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. 398, 1700–1712. doi: 10.1016/S0140-6736(21)02143-7
- Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Rev. Neurosci.* 3, 655–666. doi: 10.1038/nrn894
- Craig, A. D. (2003). Interoception: the sense of the physiological condition of the body. *Curr. Opin. Neurobiol.* 13, 500–505. doi: 10.1016/S0959-4388(03)00090-4
- Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555
- Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat. Neurosci.* 3, 1049–1056. doi: 10.1038/79871
- de Greck, M., Bölter, A. F., Lehmann, L., Ulrich, C., Stockum, E., Enzi, B., et al. (2013). Changes in brain activity of somatoform disorder patients during emotional empathy after multimodal psychodynamic psychotherapy. *Front. Hum. Neurosci.* 7, 410. doi: 10.3389/fnhum.2013.00410
- Dichter, G. S., Felder, J. N., Petty, C., Bizzell, J., Ernst, M., and Smoski, M. J. (2009). The effects of psychotherapy on neural responses to rewards in major depression. *Biol. Psychiatry* 66, 886–897. doi: 10.1016/j.biopsych.06.021
- Dichter, G. S., Felder, J. N., and Smoski, M. J. (2010). The effects of Brief Behavioral Activation Therapy for Depression on cognitive control in affective contexts: an fMRI investigation. *J. Affect. Disord.* 126, 236–244. doi: 10.1016/j.jad.03.022
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., and Fox, P. T. (2012). Activation likelihood estimation revisited. *Neuroimage*. 59, 2349–2361. doi: 10.1016/j.neuroimage.2011.09.017
- Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., et al. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *Neuroimage*. 137, 70–85. doi: 10.1016/j.neuroimage.2016.04.072
- Esposito, R., Cieri, F., di Giannantonio, M., and Tartaro, A. (2018). The role of body image and self-perception in anorexia nervosa: the neuroimaging perspective. *J. Neuropsychol.* 12, 41–52. doi: 10.1111/jnp.12106
- Fonagy, P. (2015). The effectiveness of psychodynamic psychotherapies: an update. *World Psychiatry*. 14, 137–150. doi: 10.1002/wps.20235
- Fonzo, G. A., Goodkind, M. S., Oathes, D. J., Zaiko, Y. V., Harvey, M., Peng, K. K., et al. (2017). Selective effects of psychotherapy on frontopolar cortical function in PTSD. *Am. J. Psychiatry* 174, 1175–1184. doi: 10.1176/appi.ajp.2017.1609.1073
- Fonzo, G. A., Goodkind, M. S., Oathes, D. J., Zaiko, Y. V., Harvey, M., Peng, K. K., et al. (2021). Amygdala and insula connectivity changes following

- psychotherapy for posttraumatic stress disorder: a randomized clinical trial. *Biol. Psychiatry*. 89, 857–867. doi: 10.1016/j.biopsych.11.021
- Franklin, G., Carson, A. J., and Welch, K. A. (2016). Cognitive behavioural therapy for depression: systematic review of imaging studies. *Acta Neuropsychiatr.* 28, 61–74. doi: 10.1017/neu.2015.41
- Freud, S. (1915). Instincts and their vicissitudes. *Stand. Ed. Compl. Psychol. Works Sigmund Freud*. 14, 109–140.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8
- Fu, C. H., Williams, S. C., Cleare, A. J., Scott, J., Mitterschiffthaler, M. T., Walsh, N. D., et al. (2008). Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biol. Psychiatry* 64, 505–512. doi: 10.1016/j.biopsych.2008.04.033
- Fu, C. H. Y., Steiner, H., and Costafreda, S. G. (2013). Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiol. Dis.* 52, 75–83. doi: 10.1016/j.nbd.05.008
- Furukawa, T. A., Shinohara, K., Sahker, E., Karyotaki, E., Miguel, C., Ciharova, M., et al. (2021). Initial treatment choices to achieve sustained response in major depression: a systematic review and network meta-analysis. *World Psychiatry*. 20, 387–396. doi: 10.1002/wps.20906
- Garakani, A., Murrough, J. W., Freire, R. C., Thom, R. P., Larkin, K., Buono, F. D., et al. (2020). Pharmacotherapy of anxiety disorders: current and emerging treatment options. *Front Psychiatry*. 11, 595584. doi: 10.3389/fpsy.2020.595584
- Gillies, A., and Arbutnot, G. (2000). Computational models of the basal ganglia. *Mov. Disord.* 15, 762–70. doi: 10.1002/1531-8257(200009)15:5<762::aid-mds1002>3.0.co;2-2
- Goldin, P., Ramel, W., and Gross, J. (2009). Mindfulness meditation training and self-referential processing in social anxiety disorder: behavioral and neural effects. *J. Cogn. Psychother.* 23, 242–257. doi: 10.1891/0889-8391.23.3.242
- Goldin, P. R., and Gross, J. J. (2010). Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. *Emotion* 10, 83–91. doi: 10.1037/a0018441
- Gotink, R. A., Meijboom, R., Vernooij, M. W., Smits, M., and Hunink, M. G. (2016). 8-week Mindfulness based stress reduction induces brain changes similar to traditional long-term meditation practice - a systematic review. *Brain Cogn.* 108, 32–41. doi: 10.1016/j.bandc.2016.07.001
- Graybiel, A. M., Aosaki, T., Flaherty, A. W., and Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science* 265, 1826–1831. doi: 10.1126/science.8091209
- Haut, K. M., Lim, K. O., and MacDonald, A. III. (2010). Prefrontal cortical changes following cognitive training in patients with chronic schizophrenia: effects of practice, generalization, and specificity. *Neuropsychopharmacology* 35, 1850–1859. doi: 10.1038/npp.2010.52
- Hu, S., Ide, J. S., Zhang, S., and Li, C. R. (2016). The right superior frontal gyrus and individual variation in proactive control of impulsive response. *J. Neurosci.* 36, 12688–12696. doi: 10.1523/JNEUROSCI.1175-16.2016
- Huang, X., Huang, P., Li, D., Zhang, Y., Wang, T., Mu, J., et al. (2014). Early brain changes associated with psychotherapy in major depressive disorder revealed by resting-state fMRI: evidence for the top-down regulation theory. *Int. J. Psychophysiol.* 94, 437–444. doi: 10.1016/j.ijpsycho.10.011
- Huber, D., Zimmermann, J., Henrich, G., and Klug, G. (2012). Comparison of cognitive-behaviour therapy with psychoanalytic and psychodynamic therapy for depressed patients – a three-year follow-up study. *Z. Psychosom. Med. Psychother.* 58, 299–316. doi: 10.13109/zptm.2012.58.3.299
- Huhn, M., Tardy, M., Spinelli, L. M., Kissling, W., Foerstl, H., Pitschel-Walz, G., et al. (2014). Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. *JAMA Psychiatry*. 71, 706–715. doi: 10.1001/jamapsychiatry.2014.112
- Iwabuchi, S. J., Krishnadas, R., Li, C., Auer, D. P., Radua, J., and Palaniyappan, L. (2015). Localized connectivity in depression: a meta-analysis of resting state functional imaging studies. *Neurosci. Biobehav. Rev.* 51, 77–86. doi: 10.1016/j.neubiorev.2015.01.006
- Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., and Graybiel, A. M. (1999). Building neural representations of habits. *Science* 286, 1745–1749. doi: 10.1126/science.286.5445.1745
- Jorm, A. F., Patten, S. B., Brugh, T. S., and Mojtabai, R. (2017). Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* 16, 90–99. doi: 10.1002/wps.20388
- Kandel, E. R. (1999). Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited. *Am. J. Psychiatry*. 156, 505–524. doi: 10.1176/ajp.156.4.505
- Kandel, E. R. (2016). *Reductionism in Art and Brain Science: Bridging the Two Cultures*. Columbia University Press.
- Kant, I. (1781). *Immanuel Kant's Critique of Pure Reason (Norman Kemp Smith, Trans.)*. New York, NY: St. Martin's Press (Macmillan), 1929.
- Kant, I. (2003). *Critique of Pure Reason (M. Weigelt, Trans.)*. Penguin Classics.
- Kaplan-Solms, K., and Solms, M. (2000). *Clinical Studies in Neuropsychopharmacology: Introduction to a Depth Neuropsychology*. H. Karnac (Books).
- King, A. P., Block, S. R., Sripada, R. K., Rauch, S., Giardino, N., Favorite, T., et al. (2016). Altered default mode network (DMN) resting state functional connectivity following a mindfulness-based exposure therapy for posttraumatic stress disorder (PTSD) in combat veterans of Afghanistan and Iraq. *Depress Anxiety* 33, 289–299. doi: 10.1002/da.22481
- Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., Johnson, B. T., et al. (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Med.* 5, e45. doi: 10.1371/journal.pmed.0050045
- Klein, M. (1937). “Love, guilt and reparation,” in *Love, Guilt and Reparation and Other Works, 1921–1945*. eds M. Klein (New York, NY: The Free Press), 306–343.
- Kringelbach, M. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* 6, 691–702. doi: 10.1038/nrn1747
- Kumari, V., Fannon, D., Peters, E. R., Ffytche, D. H., Sumich, A. L., Premkumar, P., et al. (2011). Neural changes following cognitive behaviour therapy for psychosis: a longitudinal study. *Brain* 134(Pt 8), 2396–2407. doi: 10.1093/brain/awr154
- Lancaster, J. L., Cykowski, M. D., McKay, D. R., Kochunov, P. V., Fox, P. T., Rogers, W., et al. (2010). Anatomical global spatial normalization. *Neuroinformatics* 8, 171–182. doi: 10.1007/s12021-010-9074-x
- Lancaster, J. L., Laird, A. R., Eickhoff, S. B., Martinez, M. J., Fox, P. M., Fox, P. T., et al. (2012). Automated regional behavioral analysis for human brain imaging. *Front. Neuroinform.* 6, e00023. doi: 10.3389/fninf.2012.00023
- Leichsenring, F., and Steinert, C. (2017). Is cognitive behavioral therapy the gold standard for psychotherapy? The need for plurality in treatment and research. *JAMA*. 318, 1323–1324. doi: 10.1001/jama.2017.13737
- Leroy, A., Vervaeke, E., Birme, P., Yger, P., Szaffarczyk, S., Lopes, R., et al. (2022). Intrusive experiences in posttraumatic stress disorder: treatment response induces changes in the directed functional connectivity of the anterior insula. *Neuroimage Clin.* 34, 102964. doi: 10.1016/j.nicl.2022.102964
- Luo, Y., Kataoka, Y., Ostinelli, E. G., Cipriani, A., and Furukawa, T. A. (2020). National prescription patterns of anti-depressants in the treatment of adults with major depression in the US between 1996 and 2015: a population representative survey based analysis. *Front. Psychiatry* 11, 35. doi: 10.3389/fpsy.2020.00035
- Ma, J. D., Wang, C. H., Huang, P., Wang, X., Shi, L. J., Li, H. F., et al. (2021). Effects of short-term cognitive-coping therapy on resting-state brain function in obsessive-compulsive disorder. *Brain Behav.* 11, e02059. doi: 10.1002/brb3.2059
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., et al. (2009). Prisma Group Reprint—preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Phys. Ther.* 89, 873–880. doi: 10.1093/ptj/89.9.873
- Menon, V., Adelman, N. E., White, C. D., Glover, G. H., and Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Hum. Brain Mapp.* 12, 131–143. doi: 10.1002/1097-0193(200103)12:3<131::aid-hbm1010>3.0.co;2-c
- Menon, V., and Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667. doi: 10.1007/s00429-010-0262-0
- Messina, D., Sambin, M., Palmieri, A., and Viviani, R. (2013). Neural correlates of psychotherapy in anxiety and depression: a meta-analysis. *PLoS ONE*. 8, e74657. doi: 10.1371/journal.pone.0074657
- Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167
- Mizumori, J. Y., Ragozzino, K. E., and Cooper, B. G. (2000). Location and head direction representation in the dorsal striatum of rats. *Psychobiology* 28, 441–462. doi: 10.3758/BF03332003
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6, e1000097. doi: 10.1371/journal.pmed.1000097

- Nabeyama, M., Nakagawa, A., Yoshiura, T., Nakao, T., Nakatani, E., Togao, O., et al. (2008). Functional MRI study of brain activation alterations in patients with obsessive-compulsive disorder after symptom improvement. *Psychiatry Res.* 163, 236–247. doi: 10.1016/j.psychres.11.001
- Nakao, T., Nakagawa, A., Yoshiura, T., Nakatani, E., Nabeyama, M., Yoshizato, C., et al. (2005). Brain activation of patients with obsessive-compulsive disorder during neuropsychological and symptom provocation tasks before and after symptom improvement: a functional magnetic resonance imaging study. *Biol. Psychiatry* 57, 901–910. doi: 10.1016/j.biopsych.12.039
- Northoff, G., and Scalabrini, A. (2021). “Project for a spatiotemporal neuroscience” - Brain and psyche share their topography and dynamic. *Front. Psychol.* 12, 717402. doi: 10.3389/fpsyg.2021.717402
- Oberhuber, M., Parker, J., Hope, T. M., Prejawa, S., Seghier, M. L., and Green, D. W. (2013). Functionally distinct contributions of the anterior and posterior putamen during sublexical and lexical reading. *Front. Hum. Neurosci.* 7, 787. doi: 10.3389/fnhum.2013.00787
- Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annu. Rev. Neurosci.* 25, 563–593. doi: 10.1146/annurev.neuro.25.112701.142937
- Pantazatos, S. P., Yttredahl, A., Rubin-Falcone, H., Kishon, R., Oquendo, M. A., John Mann, J., et al. (2020). Depression-related anterior cingulate prefrontal resting state connectivity normalizes following cognitive behavioral therapy. *Eur. Psychiatry* 63, e37. doi: 10.1192/j.eurpsy.2020.34
- Penadés, R., Pujol, N., Catalán, R., Massana, G., Rametti, G., García-Rizo, C., et al. (2013). Brain effects of cognitive remediation therapy in schizophrenia: a structural and functional neuroimaging study. *Biol. Psychiatry* 73, 1015–1023. doi: 10.1016/j.biopsych.01.017
- Qin, P., Wang, M., and Northoff, G. (2020). Linking bodily, environmental and mental states in the self-A three-level model based on a meta-analysis. *Neurosci. Biobehav. Rev.* 115, 77–95. doi: 10.1016/j.neubiorev.05.004
- Rabeyron, T. (2021). Beyond the death drive: entropy and free energy. *Int. J. Psychoanal.* 102, 878–905. doi: 10.1080/00207578.2021.1932514
- Rabeyron, T. (2022). Psychoanalytic psychotherapies and the free energy principle. *Front. Hum. Neurosci.* 16, 929940. doi: 10.3389/fnhum.2022.929940
- Radua, J., and Mataix-Cols, D. (2012). Meta-analytic methods for neuroimaging data explained. *Biol. Mood Anxiety Disord.* 2, 6. doi: 10.1186/2045-5380-2-6
- Ritchey, M., Dolcos, F., Eddington, K. M., Strauman, T. J., and Cabeza, R. (2011). Neural correlates of emotional processing in depression: changes with cognitive behavioral therapy and predictors of treatment response. *J. Psychiatr. Res.* 45, 577–587. doi: 10.1016/j.jpsychires.09.007
- Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230
- Rolls, E. T., Cheng, W., and Feng, J. (2020). The orbitofrontal cortex: reward, emotion and depression. *Brain Commun.* 2, fcaa196. doi: 10.1093/braincomms/fcaa196
- Salminen, J. K., Karlsson, H., Hietala, J., Kajander, J., Aalto, S., Markkula, J., et al. (2008). Short-term psychodynamic psychotherapy and fluoxetine in major depressive disorder: a randomized comparative study. *Psychother. Psychosom.* 77, 351–357. doi: 10.1159/000151388
- Sankar, A., Melin, A., Lorenzetti, V., Horton, P., Costafreda, S. G., Fu, C. H. Y., et al. (2018). A systematic review and meta-analysis of the neural correlates of psychological therapies in major depression. *Psychiatry Res. Neuroimaging* 279, 31–39. doi: 10.1016/j.psychres.2018.07.002
- Scalabrini, A., Mucci, C., and Northoff, G. (2018). Is our self related to personality? A neuropsychodynamic model. *Front. Hum. Neurosci.* 12, 346. doi: 10.3389/fnhum.2018.00346
- Scalabrini, A., Mucci, C., and Northoff, G. (2022). The nested hierarchy of self and its trauma: in search for a synchronic dynamic and topographical re-organization. *Front. Hum. Neurosci.* 16, 980353. doi: 10.3389/fnhum.2022.980353
- Sculpt, M. A., Fresco, D. M., Gunning, F. M., Liston, C., Seeley, S. H., García, E., et al. (2019). Changes in functional connectivity following treatment with emotion regulation therapy. *Front. Behav. Neurosci.* 13, 10. doi: 10.3389/fnbeh.2019.00010
- Shedler, J. (2010). The efficacy of psychodynamic psychotherapy. *Am. Psychol.* 65, 98–109. doi: 10.1037/a0018378
- Sheynin, J., Duval, E. R., King, A. P., Angstadt, M., Phan, K. L., Simon, N. M., et al. (2020). Associations between resting-state functional connectivity and treatment response in a randomized clinical trial for posttraumatic stress disorder. *Depress Anxiety* 37, 1037–1046. doi: 10.1002/da.23075
- Siegle, G. J., Thompson, W. K., Collier, A., Berman, S. R., Feldmiller, J., Thase, M. E., et al. (2012). Toward clinically useful neuroimaging in depression treatment: prognostic utility of subgenual cingulate activity for determining depression outcome in cognitive therapy across studies, scanners, and patient characteristics. *Arch. Gen. Psychiatry* 69, 913–924. doi: 10.1001/archgenpsychiatry.2012.65
- Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. Profile Books Limited.
- Solms, M., and Turnbull, O. (2002). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. Other Press.
- Solms, M. L. (2018). The neurobiological underpinnings of psychoanalytic theory and therapy. *Front. Behav. Neurosci.* 12, 294. doi: 10.3389/fnbeh.2018.00294
- Sözeri-Varma, G., and Karadağ, F. (2012). The biological effects of psychotherapy in major depressive disorders: a review of neuroimaging studies. *Psychology* 3, 857. doi: 10.4236/psych.2012.310129
- Steinert, C., Munder, T., Rabung, S., Hoyer, J., and Leichsenring, F. (2017). Psychodynamic therapy: as efficacious as other empirically supported treatments? A meta-analysis testing equivalence of outcomes. *Am. J. Psychiatry* 174, 943–953. doi: 10.1176/appi.ajp.2017.17010057
- Talairach, J., and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. New York, NY: Thieme Medical Publishers, Inc.
- Teichmann, M., Rosso, C., Martini, J. B., Bloch, I., Brugière, P., Duffau, H., et al. (2015). Cortical-subcortical syntax pathway linking Broca’s area and the striatum. *Hum. Brain Mapp.* 36, 2270–2283. doi: 10.1002/hbm.22769
- Thorsen, A. L., Hagland, P., Radua, J., Mataix-Cols, D., Kvale, G., Hansen, B., et al. (2018). Emotional processing in obsessive-compulsive disorder: a systematic review and meta-analysis of 25 functional neuroimaging studies. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 3, 563–571. doi: 10.1016/j.bpsc.2018.01.009
- Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., Fox, P., et al. (2012). Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. *Hum. Brain Mapp.* 33, 1–13. doi: 10.1002/hbm.21186
- Turner, E., Matthews, A., Linardatos, E., Tell, R., and Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* 358, 252–260. doi: 10.1056/NEJMs065779
- Uddin, L. Q., Kelly, A. M., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2009). Functional connectivity of default mode network components: correlation, anticorrelation, and causality. *Hum. Brain Mapp.* 30, 625–637. doi: 10.1002/hbm.20531
- Vargas-Cáceres, S., Cera, N., and Nobre, P. (2021). The impact of psychosis on sexual functioning: a systematic review. *J. Sexual Med.* 18, 457–466. doi: 10.1016/j.jsxm.2020.12.007
- Verbruggen, F., and Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends Cogn. Sci.* 12, 418–424. doi: 10.1016/j.tics.2008.07.005
- Viñas-Guasch, N., and Wu, Y. J. (2017). The role of the putamen in language: a meta-analytic connectivity modeling study. *Brain Struct. Funct.* 222, 3991–4004. doi: 10.1007/s00429-017-1450-y
- Vuper, T. C., Philippi, C. L., and Bruce, S. E. (2021). Altered resting-state functional connectivity of the default mode and central executive networks following cognitive processing therapy for PTSD. *Behav. Brain Res.* 409, 113312. doi: 10.1016/j.bbr.2021.113312
- Wakefield, S., Kellett, S., Simmonds-Buckley, M., Stockton, D., Bradbury, A., Delgadillo, J., et al. (2021). Improving access to psychological therapies (IAPT) in the United Kingdom: a systematic review and meta-analysis of 10-years of practice-based evidence. *Br. J. Clin. Psychol.* 60, 1–37. doi: 10.1111/bjc.12259
- Wells, G. A., Shea, B., O’Connell, D., Peterson, J., Welch, V., Losos, M., et al. (2022). *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality If Nonrandomized Studies in Meta-Analyses*. Available online at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm
- Wiener, S. I. (1993). Spatial and behavioral correlates of striatal neurons in rats performing a self-initiated navigation task. *J. Neurosci.* 13, 3802–3817. doi: 10.1523/JNEUROSCI.13-09-03802.1993
- Winnicott, D. W. (1949). Hate in the counter-transference. *Int. J. Psycho-Anal.* 30, 69–74.
- Wiswede, D., Taubner, S., Buchheim, A., Münte, T. F., Stasch, M., Cierpka, M., et al. (2014). Tracking functional brain changes in patients with depression under

psychodynamic psychotherapy using individualized stimuli. *PLoS ONE* 9, e109037. doi: 10.1371/journal.pone.0109037

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., et al. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* 21, 655–679. doi: 10.1016/j.euroneuro.2011.07.018

Wolters, A. F., van de Weijer, S. C. F., Leentjens, A. F. G., Duits, A. A., Jacobs, H. I. L., and Kuijf, M. L. (2019). Resting-state fMRI in Parkinson's disease patients with cognitive impairment: a meta-analysis. *Parkins. Related Disord.* 62, 16–27. doi: 10.1016/j.parkreldis.2018.12.016

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Okada, G., Kunisato, Y., et al. (2014). Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. *Soc. Cogn. Affect. Neurosci.* 9, 487–493. doi: 10.1093/scan/nst009

Yoshino, A., Okamoto, Y., Okada, G., Takamura, M., Ichikawa, N., Shibasaki, C., et al. (2018). Changes in resting-state brain networks after cognitive-behavioral therapy for chronic pain. *Psychol. Med.* 48, 1148–1156. doi: 10.1017/S0033291717002598

Yuan, M., Zhu, H., Qiu, C., Meng, Y., Zhang, Y., Ren, Z., et al. (2018). Altered regional and integrated resting-state brain activity in general social anxiety disorder patients before and after group cognitive behavior therapy. *Psychiatry Res. Neuroimaging* 272, 30–37. doi: 10.1016/j.psychnresns.2017.12.004

Yuan, M., Zhu, H., Qiu, C., Meng, Y., Zhang, Y., Shang, J., et al. (2016). Group cognitive behavioral therapy modulates the resting-state functional connectivity of amygdala-related network in patients with generalized social anxiety disorder. *BMC Psychiatry* 16, 198. doi: 10.1186/s12888-016-0904-8

Zeki, S., and Romaya, J. P. (2008). Neural correlates of hate. *PLoS ONE* 3, e3556. doi: 10.1371/journal.pone.0003556

Zhao, X. R., Chen, Z. F., Kang, C. Y., Liu, R. X., Bai, J. Y., Cao, Y. P., et al. (2019). Mindfulness-based cognitive therapy is associated with distinct resting-state neural patterns in patients with generalized anxiety disorder. *Asia Pac. Psychiatry* 11, e12368. doi: 10.1111/appy.12368

Zhao, Z., Wang, C. H., Ma, J. D., Shan, X., Shi, L. J., Wang, X., et al. (2021). Decreased left amygdala functional connectivity by cognitive-coping therapy in obsessive-compulsive disorder. *Mol. Psychiatry* 26, 6952. doi: 10.1038/s41380-021-01131-z

Zhu, X., Suarez-Jimenez, B., Lazarov, A., Helpman, L., Papini, S., Lowell, A., et al. (2018). Exposure-based therapy changes amygdala and hippocampus resting-state functional connectivity in patients with posttraumatic stress disorder. *Depress Anxiety* 35, 974–984. doi: 10.1002/da.22816

Zilcha-Mano, S., Dinger, U., McCarthy, K. S., Barrett, M. S., and Barber, J. P. (2014). Changes in well-being and quality of life in a 876 randomized trial comparing dynamic psychotherapy and pharmacotherapy for major depressive disorder. *J. Affect Disord.* 152–154, 538–542. doi: 10.1016/j.jad.2013.10.015



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Igor Douven,
Université Paris-Sorbonne, France
Nithin George,
Indian Institute of Technology
Gandhinagar, India

*CORRESPONDENCE

D. Eric Chamberlin
✉ Eric@ChamberlinNeuroscience.com

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 28 May 2022

ACCEPTED 12 December 2022

PUBLISHED 04 January 2023

CITATION

Chamberlin DE (2023) The Active
Inference Model of Coherence
Therapy.
Front. Hum. Neurosci. 16:955558.
doi: 10.3389/fnhum.2022.955558

COPYRIGHT

© 2023 Chamberlin. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

The Active Inference Model of Coherence Therapy

D. Eric Chamberlin*

Chamberlin Applied Neuroscience, Glastonbury, CT, United States

Coherence Therapy is an empirically derived experiential psychotherapy based on Psychological Constructivism. Symptoms are viewed as necessary output from an implicit model of the world. The therapist curates experiences and directs attention toward discovering the model. Rendered explicit, the model is juxtaposed with contradictory knowledge driving memory re-consolidation with resolution of the symptom. The Bayesian Brain views perception and action as inferential processes. Prior beliefs are combined in a generative model to explain the hidden causes of sensations through a process of Active Inference. Prior beliefs that are poor fits to the real world are suboptimal. Suboptimal priors with optimal inference produce Bayes Optimal Pathology with behavioral symptoms. The Active Inference Model of Coherence Therapy posits that Coherence Therapy is a dyadic act of therapist guided Active Inference that renders the (probable) hidden causes of a client's behavior conscious. The therapist's sustained attention on the goal of inference helps to overcome memory control bias against retrieval of the affectively charged suboptimal prior. Serial experiences cue memory retrieval and re-instantiation of the physiological/affective state that necessitates production of the symptom in a particular context. As this process continues there is a break in modularity with assimilation into broader networks of experience. Typically, the symptom produced by optimal inference with the suboptimal prior is experienced as unnecessary/inappropriate when taken out of the particular context. The implicit construct has been re-represented and rendered consciously accessible, by a more complex but more accurate model in which the symptom is necessary in some contexts but not others. There is an experience of agency and control in symptom creation, accompanied by the spontaneous production of context appropriate behavior. The capacity for inference has been restored. The Active Inference Model of Coherence Therapy provides a framework for Coherence Therapy as a computational process which can serve as the basis for new therapeutic interventions and experimental designs integrating biological, cognitive, behavioral, and environmental factors.

KEYWORDS

Coherence Therapy, memory control and access, Bayesian Brain, Free Energy Principle and Active Inference (FEP-AI) framework, post-freudian psychology, psychotherapy, memory representation, conscious access

Introduction to the Active Inference Model of Coherence Therapy (AIMCT)

Freud's Project for a Scientific Psychology attempted to ground Psychology in the Natural Sciences (Freud, 1895). Steeped in physiology and Thermodynamics, Freud came to view living things as dynamic energy systems. He envisioned Psychology as explainable through the physiology of the brain (Cieri and Esposito, 2019). Ultimately Freud was forced to abandon the Project due to the technical and conceptual limitations of the time. This included what he considered "the heart of the riddle," the physiology of the phenomenon of repression. In doing so he appears to have reluctantly adopted a Functionalist perspective that the mind can be studied independently of the brain (Johnson-Laird, 1983). Having abandoned pursuit of a deep connection to neurophysiology, Freud developed the meta-psychology that informs Psychoanalysis.

Recent developments have reanimated Freud's vision of understanding Psychology as the physiology of the brain (Eliasmith, 2003). The identification of resting state networks has facilitated a preliminary mapping from the psychological concepts of Psychoanalysis to the activity of the Default network (Carhart-Harris and Friston, 2010). This is part of a larger body of work known as Neuro-psychoanalysis (Solms and Turnbull, 2011; Johnson and Flores Mosri, 2016). Extending beyond Psychoanalysis, network function has been used to model the pathophysiology of Post Traumatic Stress Disorder (PTSD), and to its resolution with Eye Movement Desensitization and Reprocessing (EMDR) (Lanius et al., 2015; Chamberlin, 2019a). Viewed from this perspective the distance between the psychological and the physiological appears to diminish in what has been called Dual Aspect Monism. Briefly, "the mind and brain are emergent aspects of the same neuronal dynamics" (Hobson et al., 2021).

Another recently developed tool is the Free Energy Principle, and its realization in the Bayesian Brain (Friston and Stephan, 2007; Friston, 2009; Holmes and Nolte, 2019). The Free Energy Principle as developed by Friston et al. is an information theoretic isomorph of Statistical Thermodynamics that begins from consideration of "first principles." That is, what a living thing must do to continue living (Clark, 2013). In order to exist a living thing must resist the dissipative forces of entropy by minimizing surprising, catastrophic exchanges with the environment e.g., being a fish out of water. To accomplish this the agent must reduce its Free Energy, which is essentially the error of its predictions (Hohwy, 2013). In other words, to stay alive an agent must have a model of the world that is capable of generating predictions about sensory data, and must minimize the error of those predictions (Friston and Stephan, 2007). The Free Energy Principle through prediction

error minimization offers the possibility of explaining disparate psychological functions like perception, attention, and action with a conceptually simple mechanism that is physiologically plausible and empirically verifiable (Friston, 2010). The non-negotiable requirements of thermodynamics regarding energy and information apply at all levels of consideration e.g., from molecular to psychological (Strelnikov, 2014). This creates a common language or currency between disciplines that facilitates a deeper connection than simple associations between mind and brain (Clark, 2016). Thus the Free Energy Principle has been used to provide putative links from Psychoanalysis to neurophysiology (Hopkins, 2012, 2016; Johnson and Flores Mosri, 2016). Post-Freudian therapeutic approaches have also benefited from the explanatory power of the Free Energy Principle. For example, the Predictive Processing Model of EMDR suggests how the epistemic affordance of eye movements can be used to drive free energy minimization and belief updating in PTSD (Chamberlin, 2019b).

The Active Inference Model of Coherence Therapy attempts to leverage the explanatory power of the Free Energy Principle by offering critical analysis of an empirically validated post-Freudian psychotherapy through the lens of Bayes Optimal Pathology (Schwartenbeck et al., 2015). Application of Bayes Optimal Pathology presents two key questions for computational neuropsychology (Parr et al., 2018). The first question is "what are the prior beliefs that would have to be held to make this behavior optimal?" Next, having identified the (putative) suboptimal priors, the question becomes "what are the biological substrates of these priors?" Or, what happens in the brain? These questions will be addressed sequentially in what can overall be considered an exercise in Psychodynamic Neuroscience. In essence, using the Free Energy Principle as a bridge between mind and brain (Cieri and Esposito, 2019).

Introduction to Coherence Therapy

Coherence Therapy (CT) is empirically derived experiential psychotherapy developed by Bruce Ecker, Laurel Hulley and Robin Ticic (Ecker et al., 2012). Based on Psychological Constructivism, the emphasis of Coherence Therapy is on the individual's subjective experience of reality resulting from models built to understand the world, and to support successful action in it. To wit: "Perhaps most central to the constructivist vision is the contention that we do not passively perceive the world as it actually is, but rather shape and form what we know and experience through active, constructive mental processes" (Toomey and Ecker, 2007). And further, "...the function of forming "knowings" is to optimize the person's adaptation in the experiential world, not to accurately discern the true nature of things." With these assumptions Coherence Therapy is firmly embedded in the Constructivist tradition.

Coherence Therapy distinguishes itself from other psychotherapies with several additional assumptions. The first is that a client's presenting symptom is assumed to be a logical, "coherent" product of the "implicit knowing" or construction of reality held by the person in the present (Ecker et al., 2012). Specifically, there is an implicit "emotional truth" that makes it necessary to have or produce the symptom (Ecker and Hulley, 1996). From this perspective, despite being undesirable, the symptom is not a "dysfunction" of the brain. Per Ecker et al. "... a client's seemingly irrational, out-of-control presenting symptom is actually a sensible, cogent, orderly expression of the person's existing constructions of self and world, not a "disorder" or pathology" (Toomey and Ecker, 2007). The problem lies with the implicit model. The central task of the therapist is to facilitate an experience of the implicit "emotional truth" becoming explicit.

Following symptom identification, Discovery is the phase of Coherence Therapy which attempts to curate experiences that render the emotion and implicit model of the world that necessitates symptom production, explicit (Ecker et al., 2012). Examples of experiences include symptom deprivation, sentence completion, role play, imaginal interaction techniques etc (Ecker, 2016). The Principle of Immediate Accessibility asserts that the implicit model can be known rapidly (from one to several sessions depending on the underlying complexity of the material) (Ecker and Hulley, 1996; Ecker, 2018a). This stands in contrast to conventional wisdom in psychotherapy about the difficulty of reliably accessing unconscious material swiftly, and is based on Ecker's research developing the experiential techniques that facilitate this process. It appears that the assumed coherence or logical necessity of the symptom given the implicit model facilitates efficient elucidation of the model (Ecker et al., 2012).

An important element of the retrieval of the implicit model or schema in the Discovery process is the sustained application of attention to the task by both parties, driven by the therapist (Ecker and Hulley, 1996). From the perspective of the client, the symptom is problematic and the idea that the symptom "makes sense" at some level is deeply counter-intuitive. Thus, the success of the Discovery process is contingent on the therapist's commitment to the assumption of coherence to overcome the intuitive bias of the client and sustain focus on Discovery (Ecker et al., 2012). An additional challenge is that the implicit model is typically formed early in life in a situation with emotional urgency, often leading to repression or exclusion from consciousness of the emotionally charged memory (Ecker, 2015).

Following Discovery, the implicit model has been rendered explicit and the Integration phase ensues. Integration is intended to facilitate incorporation of the now explicit model into everyday awareness in preparation for the final phase of Juxtaposition (Ecker et al., 2012). In Juxtaposition the client is ushered into an experience of simultaneous awareness

of the symptom necessitating model, and personally held contradictory living knowledge. This juxtaposition results in effortless cessation of the symptom going forward, having presumably driven prediction error mediated memory reconsolidation (Ecker et al., 2012). Typically, after having been rendered explicit the symptom is understood as being necessary in some contexts, e.g., past personal experience, but not in the present context (Ecker and Hulley, 1996). The result is an experience of agency in symptom production, and spontaneous context appropriate behavior thereafter.

Discovery as the essential activity of Coherence Therapy

Thus far Ecker et al. have focused on the role of the Juxtaposition phase and presumed memory re-consolidation as being the core active element of this effective therapy (Ecker, 2018a). In contrast, this review will focus on the Discovery phase and its role in the therapeutic process. The reason for this is that in practice, successful Discovery of the precise symptom necessitating schema results in immediate and enduring cessation of the symptom in more than half of clients sampled (Ecker, 2018b; Ecker and Bridges, 2020). In these cases, it appears that when the symptom necessitating schema is first apprehended clearly in consciousness, the client is immediately confronted with contradictory knowledge that often leaves them bemused e.g., "When I think about it, that's ridiculous" (needing to produce the symptom in the current context) (Ecker et al., 2012). Thus it appears that the Discovery phase is the only necessary and (often) sufficient phase of the process. While Integration and Juxtaposition likely ensure consistent, comprehensive resolution, something crucial appears to be happening in Discovery.

The Bayesian Brain

The Bayesian Brain views perception and action as inferential processes (Doya et al., 2006; Parr et al., 2018). Recognizing that the brain does not have direct access to the world it must infer the hidden causes of the sensations it experiences. For example, light does not enter the brain, the inside of the skull is dark. Upon striking the retina, the energy of light is transformed into the firing of neurons. From the resulting patterns of neuronal firing the brain must infer, or guess, the hidden cause in the world eliciting that pattern (Hohwy, 2013). The process of inference is presumed to be accomplished by physiologically plausible computations employing Bayesian logic. The prior probability (the "prior") is the probability of a cause before consideration of the particular sensory data. The likelihood is the conditional probability of the sensory data given the cause. Precision reflects confidence

in prior beliefs and in sensory data, varying how much weight should be given to each in computations. Taken together the “prior” and the “likelihood” constitute a “generative model” i.e., a model of how the sensation is generated by a hidden cause in the world. This constitutes the brain’s “best guess” (as a probability) regarding the current state of the world.

The Active Inference of embodied, embedded agents

Recognizing that inference occurs in an embodied agent, who is embedded in a particular environment in the world, it becomes apparent that it is an active process. In order to more accurately infer the causes of the sensations it experiences, the agent must act on the world and run experiments to test hypotheses about the hidden causes (Friston K. et al., 2012). For example, “what happens if I look over there, or push on this?” (Parr and Friston, 2017). The sensory results of such experiments provide data that is then used to revise current hypotheses and suggest new experiments. The best experiments resolve the most uncertainty, that is, provide the most information (Friston et al., 2016; Clark, 2018). For example looking toward the source of a sound is likely to provide more information about its cause than looking away from it. As this process of hypothesis generation, testing, revision, etc., proceeds, inferences about the current state of the world become more accurate. Formally, model evidence increases. This suggests that humans perceiving and acting in the world behave essentially as scientists. That is, from the perspective of Active Inference they employ the core elements of the Scientific Method as they learn about the world. This assertion is supported by a rich body of empiric work in childhood cognitive development (Gopnik and Wellman, 2012).

Bayes Optimal Pathology

Human behavior is not always adaptive. How is this understood from a Bayesian perspective? The results of the Complete Class Theorem suggest that for any behavior, there is a prior belief that makes that behavior “Bayes Optimal” (Wald, 1947). This includes non-adaptive or pathological behavior. In other words, the inferential process itself may be intact i.e., “Bayes Optimal.” However, starting with a suboptimal prior, the behavior resulting from intact inference may be “pathological” (Schwartenbeck et al., 2015). By analogy, despite sound logic, if one starts with questionable assumptions the conclusion may be inappropriate. The result is Bayes Optimal Pathology. This suggests that a pathological behavior “makes sense” if we know the suboptimal prior belief it is based upon. It has been suggested that the identification of the suboptimal priors underlying pathological behavior is a promising clinical

approach (Parr et al., 2018). If so, the question becomes “what are the prior beliefs that would need to be held to make a behavior Bayes Optimal”?

Consilience of Coherence Therapy and Bayes Optimal Pathology

Coherence Therapy assumes that behavior, including pathology is always “coherent,” that is “makes sense” and is logically consistent if one identifies the underlying “emotional truth” or schema on which it is based. In fact, the underlying schema necessitates the behavior, including symptom production. This is analogous to the notion of Bayes Optimal Pathology where a suboptimal prior belief (=“emotional truth”) with intact Bayesian inference, gives rise to pathological behavior (a symptom). There is not necessarily “dysfunction” in the physiology of the brain (Parr et al., 2018). As suggested previously the question becomes “what are the prior beliefs that would need to be held to make the behavior Bayes Optimal?” The analogous question in Coherence Therapy is, “In what way does your (undesirable) behavior/symptom make sense?” The process of Discovery is essentially to curate experiences that lead to identification of the suboptimal prior in this particular client. This can be thought of as an act of meta-Bayesian inference in which the therapist is making inferences about the subject’s inferential process (Daunizeau et al., 2010).

Discovery effects a representational redescription

The Discovery Phase of Coherence Therapy begins with a symptom that the subject has identified as problematic or undesirable. The shared initial goal as framed by the therapist is to identify the hidden cause of the client’s behavior, characterized as an implicit “emotional truth” or schema that necessitates symptom production. “The therapeutic target of change is a given symptom’s underlying emotional learning which is a schema, or mental model, that was learned long ago in an emotionally intense experience and has been controlling behavior and or state of mind from outside of awareness” (Ecker and Bridges, 2020). Of note is that the schema is not conscious despite its significant, problematic influence. Ecker et al. characterize finding this schema as “retrieval” of the schema. The Active Inference Model of Coherence Therapy (AIMCT) diverges significantly from this perspective viewing identification of the “symptom necessitating schema” as being the product of coordinated Active Inference that while often retrieving a relevant memory, always acts to re-represent the target schema at a higher hierarchical level. This process takes time as the back and forth exchanges of therapist and

client establish the features of the representation. This new representation is consciously accessible and is available for verbal expression and different types of cognitive processing. This includes action (policy) selection. In addition, it is embedded in context and manifests agency. In other words, the essence of the Discovery phase of Coherence Therapy is knowledge construction through Active Inference. In the context of the Bayesian Brain applied to psychotherapy this type of knowledge construction has been called Structure Learning. It refers to “learning the repertoire or narratives that constitute our prior beliefs—or hypotheses—about how our world works, and how these might be influenced therapeutically” (Holmes and Nolte, 2019). Empiric simulations suggest that “structure learning can be seen as an emergent property of active inference (and learning) under generative models with “spare capacity”; where spare or uncommitted capacity is used to expand the repertoire of representations” (Smith et al., 2020).

Convergence of Coherence Therapy and Psychoanalysis

Focusing on the notion of “representation” Hopkins synthesized the deep convergence of Freudian Metapsychology and the Bayesian Brain (Hopkins, 2012). Freud viewed symptoms as occurring in the setting of failures of First Person Authority (agency) due to emotionally conflicting representations held by the patient (Freud et al., 1961). The challenge of acting while holding conflicting representations was managed by suppression of one representation according to Freud. As in the Bayesian Brain, there can only be one best (highest probability) representation at a time. For example, in the binocular rivalry face-house image, the brain chooses the most likely representation at that moment and sees a face or a house, but not both. One representation is utilized while the other is suppressed (for the moment) (Hohwy, 2013). From Freud’s observations, however, the suppressed representation may exert effects on behavior that manifested as symptoms, without awareness of the role of the suppressed representation (This is consistent with continued evidence of the suppressed representation at lower cortical levels in binocular rivalry (Zhang et al., 2011)). Utilizing observations of behavior and free association, the psychoanalyst infers regarding the “hidden cause” (suppressed representation) of the symptom delivering a verbal interpretation to the client that attempts to make this suppressed representation explicit and consciously accessible so that the conflict, and therefore the symptom can be resolved.

Although not its stated intent, the Discovery phase of CT often elicits suppressed memories as it “retrieves” the symptom necessitating schema, that is usually grounded in an emotionally intense childhood experience (Ecker, 2018a). Once explicit, the conflict and symptom are resolved with the emergence of agency, or what Freud called First Person Authority.

The challenge of never represented “inchoate forces”

In addition to representations that were suppressed and therefore unavailable to consciousness, Freud also identified “inchoate forces” that had never been represented and yet had causal (behavioral) efficacy (Levine, 2012). Never having been consciously represented these implicit influences pose a significant problem for Psychoanalysis. The challenge of representing these influences verbally so that they can be thought, and processed further, has been called “the work of Psychic Figurability” (Botella and Botella, 2004). Because of the amorphous nature of these unconscious influences a significant concern with this work lies in ensuring that the jointly rendered explicit representation is the product of the client’s, and not the analyst’s mind, as much as possible (Levine, 2012). A further technical challenge is the required level of therapist activity which deviates significantly from the core Psychoanalytic practice of observation.

The AIMCT argues that Coherence Therapy addresses the challenge not only of rendering suppressed representations explicit, but also making those that have never been represented, explicit. It achieves this through coordinated active inference that utilizes therapist observation to make inferences that are utilized not as interpretations, but to design experiments that engender visceral experience and prediction error constrained disambiguation of hypotheses regarding the hidden causes of the client’s behavior. In the process a representational redescription is effected. Because therapist inference is only used to design disambiguating experiments, therapist bias and “contamination” are (hopefully) minimized. “To do Coherence Therapy is to say and do nothing designed to directly counteract, overcome or prevent the symptom, and to do no interpreting” (Ecker, 2016). “Better narrative, clever reframes, positive beliefs and other “counteractive” interventions are precluded” (Ecker et al., 2012). This helps to operationalize the constructivist assumption of “the client as the expert on their life,” with the therapist coming from a “not knowing” but actively curious position (Anderson, 1992). Requiring therapist inferences to be utilized only through experimental design helps to create a “firewall” that attempts to keep therapist models separate from the client’s (evolving) model. It likely contributes to the striking claim that “any competent (Coherence) Therapist should discover the same set of constructs” (Ecker and Hulley, 1996).

Divergence of Coherence Therapy from Psychoanalysis

Like Psychoanalysis, Coherence Therapy makes extensive use of observation. However, unlike Psychoanalysis it is an intrinsically active therapy that uses experiments to create

experiences and disambiguate competing hypotheses (Ecker, 2016). From the AIMCT perspective it is an iterative process where the therapist inverts his evolving generative model of the client's (hidden) generative model to specify the subject's likely suboptimal priors which are then subject to experimentation. This strategy is consistent with the suggestion that identification of possible suboptimal priors for experimentation is a starting point for phenotyping in Computational Psychiatry (Schwartenbeck and Friston, 2016). The core divergence of these two therapies is captured in the rough analogy: Psychoanalysis is to Perceptual Inference as CT is to Active Inference. That is, passive observation is replaced by active exploration and experimentation. While Perceptual Inference operates on latent states that cannot be controlled, active inference operates on states that can be controlled e.g., actively guiding another's experience to gain a (mutual) understanding of why someone does something (Frith, 2012). Understanding how such representations may become explicit requires a deeper look at the representational redescription Model.

Representational redescription begins with implicit representations

The model of representational redescription was developed to explain observations of cognitive development in children as they acquire language and solve problems (Karmiloff-Smith, 1979). As children learn they appear to pass through several stereotyped phases beginning with implicit representations that reflect knowledge of a procedure. That is, they are able to produce a particular output in response to particular external stimuli. However, while their behavior suggests that they know "what to do" in a given situation, they are unable to describe what they are doing, or why. At this level of implicit representation, "it is knowledge in the system, but it is not yet knowledge to the system" (Clark and Karmiloff-Smith, 1993). As a result, the knowledge that is embedded in the procedure is not available to other processors or the system as a whole. In that sense the implicit knowledge is "modular" (Karmiloff-Smith, 1994).

Implicit representations are serially re-described and rendered explicit

The representational redescription model posits a progressive "explicitization" of the initial implicit procedural representation (Karmiloff-Smith, 1986). Based on empiric observations, the essential point is that the implicit representation is not rendered explicit as a result of simple "reading" by a conscious or semantic processor. Explicitization is thought to be a constructive process that takes time and often

requires some type of "scaffolding" or "cognitive processing prop" to be effected. Successive iterations generate increasingly more likely representations as hypotheses are tested and revised (Gopnik and Wellman, 2012). An analogy is drawn to scientific discovery with the accretion of knowledge and progressive re-description resulting in a flexible knowledge structure that can be utilized in different ways. Recent work on the optimization of knowledge construction sheds light on the products of this process suggesting that the most useful schemas are composed of a balance between generalized semantic and detailed episodic memories (van Kesteren and Meeter, 2020). The AIMCT posits the Discovery Phase of Coherence Therapy is a process of progressive representational redescription that yields a relatively optimized schema with a balance of semantic generalization and episodic detail that typically renders the presenting symptom unnecessary, or inappropriate in the current context.

Learning to be conscious

The concept of representational redescription has been incorporated into recent work that sheds light on the dynamics of the process of redescription. The Self-Organizing Meta-representational Account of consciousness (SOMA) suggests that consciousness is something the brain learns to do (Cleeremans et al., 2020). Driven by the need to manage the consequences of its actions on the world in order to ensure survival, the brain develops a model of itself. Starting with neuronal activity which is intrinsically non-conscious, the brain learns how to execute an action in response to certain stimuli without awareness of what it is doing. For example, a young child may have a sensation of fear and complain of a (non-existent) stomach ache. This results in staying home from school and avoiding a feared experience. There is not necessarily any awareness of this "procedure" or why it is executed, it just works. The uncomfortable sensation of fear goes away. According to SOMA this type of procedure may operate automatically without awareness, indefinitely as long as it is adaptive. However, if the consequences of the behavior become sufficiently problematic e.g., sensing irritability from an inconvenienced parent, the procedure may no longer be adaptive. Then the brain is doing something in its exchanges with the world that has unexpected, undesirable consequences. In other words, the brain's activity is resulting in excessive uncertainty, and in preferred states not being realized. The brain may attempt to reduce uncertainty, and gain increased control by creating a model to explain the causes of the unexpected consequences. It does this by creating a representation of its own activity. The "procedure" with its initial sensation of fear, the act of verbalizing "stomach ache," and the effects become objects for representation. "I feel fear, say I have a stomach ache, miss school, the fear goes away, and mom is irritated." The brain

has made its activity an object for itself to be represented. It is creating “a theory about itself” and in so doing there is the emergence of phenomenal consciousness, agency, and control (Cleeremans et al., 2020).

The emergence of a self with agency

In the process of representing the brain’s activity to itself, the notion of a “self” necessarily emerges. To have a sensation, requires that there is an experiencer of the sensation, that is, a “self” that is having an experience (Fleming, 2020). And if an action is taken, there needs to be an agent taking that action. Thus the act of representation results in conscious awareness of a self with agency. Because this self can take action or not, there is the possibility of control.

The process of re-representing lower hierarchical levels can continue as needed to reduce uncertainty and increase control. For example, the child who has represented the original procedure may say “I was scared and said I had a stomach ache, missed a lot of school and mom got irritated” as an early representation. This may evolve into a higher level representation that is both more detailed, and more general e.g., “I was scared because I had a math test on those days, and I had failed one early in the year. When I’m scared, I avoid things.” And then a higher level “I was scared and avoided math tests because I really want to be an engineer, like my mother. So in order to do that I have to try to stop avoiding.” Of note is that one can easily imagine how this iterative re-representation could be facilitated by attentive, interested others who help “connect the dots” of self-experience e.g., parents, teachers etc.

There are several important things to note about this process. Learning to be conscious through representational redescription takes time and involves processing different types of information (perceptual, episodic memory, general knowledge etc.). This implies that the modularity of the initial implicit knowledge as noted by Karmiloff-Smith has been broken, consistent with the Global Neuronal Workspace Hypothesis. Per Dehaene et al. “For the GNW, consciousness serves a function: it evolved to break the modularity of non-conscious processing and broadcast information to a brain-wide network that makes information globally available for report (motor or verbal) and post-perceptual cognitive processing like working memory or decision making (Panagiotaropoulos et al., 2020). Also of note is that there is a progression toward models of increasing temporal depth driven by the need to reduce uncertainty and to realize preferred states. With the initial procedure there is no representation of consequences or future states e.g., mom getting irritated over time. As the re-representation process continues with additional iterations and elaboration of hierarchical levels, the expected consequences of actions are represented over progressively greater intervals of time. These inferences about the future e.g., what happens

if I do this, or if I do that, are the province of Expected Free Energy.

Models of the future-expected free energy

In the Bayesian Brain the process of belief updating is driven by the imperative to minimize the discrepancy between the prior belief or model and the observed data in order to survive (Hohwy, 2013). Too great a discrepancy or divergence between the belief and observation may be incompatible with life e.g., a fish that “expects” to be in water observing that it is not. This divergence or surprise cannot be computed directly, however, variational free energy or prediction error can be computed and serves as a bound on surprise i.e., Free energy is always greater than surprise (surprise or surprisal is used here as a formal construct of information theory, not in its colloquial sense.). Therefore, in active Inference the drive is to minimize surprise by minimizing (average) Free Energy (Friston, 2009). In the moment this can be accomplished by updating the model to bring it in to accord with the observations, or by taking action and sampling the world to obtain different observations that are closer to those that are expected. Either approach may minimize the free energy or prediction error.

Given that an agent must minimize Free Energy over time, taking action in the world entails additional challenges. Taking action implies choice i.e., to do this or that. And each choice has consequences that will unfold in the future. Therefore, in order to minimize Free Energy an agent must be able to model different counterfactual scenarios and their associated “expected free energy” (Parr et al., 2022). Such imagined sequences of actions are “policies,” and choosing between policies based on their expected free energy is Bayesian Model Selection. In generating these scenarios of possible action sequences and their consequences, one is creating “Memories of the Future” (Ingvar, 1985). This function of imagining and planning actions is thought to depend on the on the prefrontal cortex (Schacter et al., 2012; Fuster and Bressler, 2015). As such there are implications for this capacity as a function of stress and of development (Arnsten et al., 2015).

Intractable complexity as mental disorder

Using Ainsworth’s Strange Situation Hopkins illustrates the concept of computational complexity as Mental Disorder (Hopkins, 2016). Placed in a room with mom and toys, infants happily explore and play. Then mom leaves the room and the infant is left with a stranger experiencing unexpected uncertainty. Upon mom’s return, the unsettled, surprised infant needs to minimize uncertainty. Attachment

to mom prompts a kinematic trajectory of approach. Anger toward mom for abandonment prompts a trajectory of avoidance. Securely attached infants approach mom, express anger, accept comforting, settle quickly, and resume play. Those with disorganized attachment are unable to manage the computational complexity, and enact an incoherent kinematic trajectory with elements of approach and avoidance. In other words, when faced with different possible courses of action, each with its own uncertainty, the computations necessary to select a single, best course of action “seize up.” The resulting behavior reflects the inability to choose, and they remain in an unsettled state for a prolonged time. Per Hopkins: “The complexity of affordance competition and action selection in this basic social decision is so magnified by conflict as to render the required computations too emotionally complex for their generative models to manage.” Furthermore “the conflicts and traumas that Freud thought responsible for recourse to phantasy/virtual reality in mental disorder should be seen as forms of neurocomputational complexity, and that mental disorder is the product of such complexity together with the mechanisms that have evolved to reduce it.”

Self-knowledge as inference

Clinical presentation often takes the form of “something is happening to me and I want it to stop.” As observed by Freud there is typically a failure of First Person Authority, or absence of agency. It is happening “to me.” The therapist’s assumption in Coherence Therapy is that the client is producing the symptom because it is necessary according to his current construction of reality (Ecker, 2016). The accumulation of evidence to support this belief requires exploration and development of the client’s self-knowledge.

From the perspective of Active Inference, insight or explaining one’s actions can be regarded as the product of inference (Parr and Pezzulo, 2021). Attempting to explain one’s behavior retrospectively requires comparison of different policies (sequences of actions) and their expected consequences. Based on the observed consequences one can select the policy that best explains the data. “I did that because I believed that this would happen.” This is analogous to *post hoc* model selection in science where experimental outcomes are used to disambiguate competing hypotheses yielding inference to the best explanation. Agent based simulations using this framework have produced some interesting results.

Parr and Pezzulo’s findings support their hypothesis that insight is confabulation that is constrained by data to a greater or lesser degree, depending on the quality of the data used to formulate the explanation (Parr and Pezzulo, 2021). Furthermore they observed the emergence of replay, an analog of episodic memory which is both declarative, and embedded

in a context. Taken together these findings suggest that the quality of insight resulting from inference varies widely and improves with episodic declarative memories that are embedded in context.

With these observations in mind the Coherence Therapist begins Discovery with the client having essentially no insight. “This symptom is happening to me for no clear reason.” Therefore the self-knowledge inferential process begins with a therapist assumption that attempts to increase the client’s hypothesis space to include the possibility that the symptom is logical and that there is a clear (hidden) cause that involves self. In other words, a new hypothesis needs to be added to the subject’s model in order to adequately account for current observations (Smith et al., 2020). The client is invited to enter and explore the space, and a Duet for One ensues (Friston and Frith, 2015). This leads toward synchronization of the therapist and client model of symptom causality, driven by the high precision (confidence) of the therapist’s belief that the symptom is a logical product of the client’s self, driven by a particular construction of reality. This is a crucial step in repair of the client’s capacity for inference on this issue.

The process of Discovery employs a well-defined methodology, but is not prescriptive and does not have a protocol *per se* (Ecker, 2016). Rather the guiding imperative is Active Inference to the best explanation. The principal question is “What model causes production of this symptom?” All actions taken by the therapist including asking questions, suggesting experiments, clarifying responses, reflecting back observations of behavior etc., are designed to create epistemic affordances that reduce uncertainty and answer this principal question. Bayesian model selection is repeatedly enacted as the therapist considers and chooses which policies (experiments) will provide the greatest information gain and maximize reduction in uncertainty (FitzGerald et al., 2014). This process is analogous to the Scientific Method as experiments are chosen for the highest likelihood of disambiguating between competing hypotheses. The absence of a protocol in the crucial Discovery phase of CT may seem surprising, however, is supported by recent Active Inference simulations of linguistic dialog. To wit: “... the purposeful, inquisitive and abductive behaviors are all emergent properties of minimizing (expected) free energy. In other words, there is no need to handcraft any rules. . .” (Friston et al., 2020).

Consider a commonly employed experiment in Discovery. “Imagine being in the situation where the symptom typically occurs, without the symptom” (Ecker, 2016). *A priori* this is a high yield experiment as it offers an epistemic affordance as well as potential pragmatic (therapeutic) value. Recent empiric work in Endogenously Generated Emotion (EGE) suggests that simulation of a novel experience activates the Salience Network (correlating with subjective core affect) and the Default Mode Network (correlating

with the generation of representations) coordinated by the Frontoparietal Control Network. This results in an affectively charged experience of something that has never occurred (Engen et al., 2017). With memory of actual events being the substrate of simulation (Schacter et al., 2007; Buckner et al., 2008), there is a sleight of hand in this Discovery experiment in that the client is potentially guided to remembering and re-experiencing a repressed memory. The AIMCT posits that memory control bias against recall and reexperiencing of the associated physiological state, is a common cause of impaired inference that precludes representational redescription, experience of agency and contextual embedding. Policy selection in the moment of a childhood, emotionally urgent situation is unavailable for revision and optimization. As a knowledge construct, the schema lacks the desirable attributes of context or episodic detail and functions as a suboptimal prior giving rise to Bayes Optimal Pathology.

What does the Active Inference of Coherence Therapy look like?

Lacking an explicit protocol, the Discovery Phase of CT is challenging to describe. A useful analogy for the process is that of a Forensic Sketch Artist. Beginning with a sparse description from a witness, the artist renders an image or “prediction” of what the suspect looks like. The witness gives feedback on the image e.g., “the nose is too large and the lips are too thin.” The artist uses this feedback or prediction error to revise the image. Successive iterations of this process minimize the prediction error until a satisfactory, detailed representation is obtained. In a similar fashion the Coherence Therapist collects verbal reports, observations of somatic behavior, associations, results from experiments etc., and sketches the emerging model that is then subject to error correction by the client. While clients typically have great difficulty in articulating what the model is initially, they have little difficulty in identifying what is wrong with a proposed model, thus driving effective revision by the client-therapist system. This is consistent with the constructivist notion that in order to define what something is, one must define what it is not (Kelly, 1991). And as the emerging model approaches the client’s implicitly held model, there is robust affective resonance. “That’s it exactly!... that’s how it works!... my body is vibrating”... “I can’t help but smile” etc. (Ecker et al., 2012). Such affective resonance with involuntary activation of facial musculature has been captured in the clinical concepts of “limbic music” and “the face of fluency” (Murray, 1992; Topolinski et al., 2009). These responses are consistent with the reward of reduced uncertainty and structure learning that leads to “a ha!” or “eureka” moments of insight (Friston K. J. et al., 2017; Friston et al., 2020).

Discovery resembles simulated annealing

The Active Inference of Discovery as described above is essentially a search for the implicit model driving symptom production. The goal is to find the model with the greatest evidence.

Model evidence can be decomposed into accuracy minus complexity. Therefore maximizing model evidence requires attention not only to accuracy, but to complexity as well. There is a cost for complexity. Overly complex models are not useful. The Discovery process, roughly captured in the Forensic Sketch Artist analogy above, progressively optimizes model evidence with shifts in hypotheses that capture and lose accuracy, and increase and decrease complexity. Essentially this is the application of Occam’s razor, or the principle of parsimony. As attributed to Einstein, “everything should be made as simple as possible, but not simpler.” For example an overly parameterized model will be rejected by the client when it doesn’t fit a particular instance that comes to mind i.e., it doesn’t adequately generalize. As a result, parameters will be pruned and the model subject to re-testing. The more iterations, the more likely the global minima of free energy is found. This process is analogous to simulated annealing in computer science where broad “high temperature” searches gradually give way to more narrow “low temperature” searches as the optimal solution is approached (Gopnik et al., 2017). In this process the therapist plays an important role in regulating the “temperature” of the search, beginning by encouraging broad exploration in an apparently empty hypothesis space- “I don’t see how this symptom could make sense, but okay, I’ll try. ...”

Toward a process theory-how might it work in the brain?

Application of Bayes Optimal Pathology presents two key questions for computational neuropsychology (Parr et al., 2018). The first question is “what are the prior beliefs that would have to be held to make this behavior optimal?” The AIMCT has argued that a core activity of CT lies in answering this question through active inference and representational redescription that leads to Structure Learning. Having identified the (putative) suboptimal priors, the next question is “what are the biological substrates of these priors”?

Active Inference provides a principled framework for understanding behavior. As with any normative theory, its ultimate utility depends on how well it explains empiric observations and generates testable predictions (Bowers and Davis, 2012). Specifically, this means generating hypotheses grounded in physiologically plausible mechanisms that are answerable to empiric data, so called process theories (Parr et al., 2022). Process theories attempt to offer an answer of

how the phenomenon in question actually works in the brain with specific references to anatomy and physiology. Recent work offers preliminary support for some process theories suggested by Active Inference (Friston K. et al., 2017). What follows is a preliminary sketch of a process theory regarding the genesis of suboptimal priors, including the apparent compromise of the inference process that precludes optimization of these priors. How might people get stuck with beliefs that are resistant to change and cause symptoms? What structures and mechanisms might be involved?

Memory suppression regulates affect

Freud postulated that painful, unwanted memories could be excluded from awareness *via* a process he called Repression (Freud et al., 1961). Utilizing fMRI Anderson et al. confirmed the existence of an “active forgetting process” he called suppression, that established a framework for the study of “motivated forgetting” (Anderson et al., 2004). To remember entails some degree of reactivation of the physiology, and re-experiencing of the associated affect (Daselaar et al., 2008; Danker and Anderson, 2010). Memory is thus an important source of endogenously generated emotion (Engen et al., 2017). As such it has been suggested that memory control is an important mechanism of affective regulation (Benoit et al., 2014; Catarino et al., 2015; Streb et al., 2016; Engen and Anderson, 2018; Mary et al., 2020). However, while suppressing retrieval of a memory may help to regulate affect, there appear to be costs as well (Smith et al., 2016).

Memory suppression interferes with representational redescription

Suppression of memory retrieval is thought to utilize the right dorsolateral prefrontal cortex to dynamically inhibit hippocampal activity while simultaneously inhibiting extra-hippocampal regions supporting the retrieval of affect (amygdala) and scene features (sensory cortex) (Benoit and Anderson, 2012; Benoit et al., 2014; Depue et al., 2015; Gagnepain et al., 2017). The parietal cortex is believed to be a key intermediary in episodic retrieval suppression, as it is in retrieval (Shimamura, 2011; Paz-Alonso et al., 2013). Suppression of the retrieval process thus disrupts the reconstruction of the episodic memory by inhibiting the hippocampus, amygdala and sensory cortex. There is evidence to suggest that retrieval suppression with reduction in hippocampal activity may result in a “virtual lesion” that leaves an “amnesic shadow” beyond the repressed memory extending to memories that are close in time to the suppression (Hulbert et al., 2016). It also appears to render the related information unavailable for further processing e.g., representational redescription and inference.

This is consistent with Freud’s notion of a repressed memory that maintains causal efficacy driving symptom production outside of conscious awareness.

Recruiting motivation to recall a repressed memory

Typically forged in childhood in an emotionally urgent situation the “emotional truth” or implicit schema is often associated with a painful memory that has been suppressed. Accessing the memory is avoided because it is painful (Ecker, 2015). However, the presenting symptom is also problematic, if not painful. When placed in the framework of the AIMCT, the discomfort of the symptom drives exploration leading toward the retrieval of the suppressed memory and associated schema. In this context the suppressed memory is not simply a painful (re-) experience to be avoided, but a potential epistemic affordance with pragmatic utility. Retrieval of the experience may facilitate representational redescription of the implicit schema with the emergence of agency and therefore control over the symptom. This shift in memory control reflects a crucial aspect of Active Inference and the Bayesian Brain: we are in charge of selecting the data we sample in trying to adapt to our world (Pezzulo et al., 2018). And the choice of what memories are sampled appears to play a significant role in the pathogenesis of suboptimal priors. Coherence Therapy appears to recruit motivation to retrieve the painful suppressed memory thus making a different choice that leads to repair of the inference process.

Motivation drives the approach or avoidance of knowing

Recently the theory of epistemic motivation has been integrated with active inference dissolving the dichotomy between motivation and cognition (Kruglanski et al., 2020). As the quintessential mechanism of knowledge construction, inference is argued to be suffused with motivation (Pezzulo et al., 2018). The essence of this idea is that motivational bias is present in the choice of policies pursued for epistemic affordances. Sometimes we want to know e.g., “I think I did well so I’m going to try to find out what grade I received.” This leads to behavior that provides epistemic affordance e.g., going to the classroom where the grades are posted. At other times we don’t want to know e.g., “I think I failed so I’m going to get something to eat.” In this case a policy is chosen with no apparent epistemic affordance regarding the grade. More accurately, ignorance is pursued as the desired state. Of note is that the policy also reflects motivation to realize a “not hungry, satisfied state.” The epistemic affordance pursued is regarding food, not grades. From this perspective motivation plays a

crucial role in epistemic foraging and therefore the knowledge construction resulting from active inference.

Utilizing the Structural Model of neocortical organization (Barbas and Rempel-Clover, 1997) the Adaptive Bayes Process Model suggests how relatively undifferentiated limbic cortex can influence highly differentiated hetero-modal cortex thus infusing motivation into the computations of the neocortex to facilitate Motivated Control (Tucker and Luu, 2021). Limbic and subcortical homeostatic control provide affectively charged expectancies which are incorporated to facilitate allostasis, the motive control of expectancies for future events. With this, the Adaptive Bayes Process Model suggests a neuropsychology to explain how motivation can be “baked in” to cognition and the Bayesian Brain.

Returning to Discovery in CT, the therapist’s high precision (confidence) belief that an implicit model necessitating production of the symptom can be found and will be useful provides motivation for active inference for both parties. To wit: “It emerged that the key condition for brief deep work was the therapist’s conviction that the unconscious constructs generating the client’s problem are immediately accessible and changeable from the start of therapy” (Ecker and Hulley, 1996). Recent empiric work suggests that such “...optimistic expectancies involve particularly strong predictions of reward causing automatic guidance of attention to reward...” (Singh et al., 2020). This is consistent with the idea that motivation influences epistemic foraging from the start (Kruglanski et al., 2020). In addition to the therapist’s optimism, the client’s subjective distress regarding the symptom provides motivation. Potentially these jointly held motivations can prevail over the drive to suppress a relevant memory. If so, how this might work in the brain.

Attention drives low confidence memory search

The question “in what way does the symptom make sense for you?” invites a goal directed memory search. The Attention to Memory Model posits that the Dorsal Parietal Cortex is associated with the allocation of attentional resources to memory retrieval according to the goals of the person (top-down attention) (Cabeza et al., 2008). The dorsal parietal cortex is thought to modulate memory retrieval activity in the medial temporal lobe while the ventral parietal cortex detects when relevant memories have been retrieved, prompting a shift in attention to the retrieved contents (bottom-up attention) (Corbetta et al., 2008). The dorsal parietal cortex is particularly important in effortful, “attention demanding low confidence searches” where the rememberer is unsure if they know something (“vaguely familiar”) (Ciaramelli et al., 2020). If memory retrieval is successful the ventral parietal cortex will capture attention to

the MTL output with the experience of “recollection” with episodic details.

The challenge of a successful attention demanding low confidence search is reflected by the assertion that directing attention toward the coherence of the symptom (and its implicit model) is the central activity in Coherence Therapy (Ecker and Toomey, 2008). Therapist attention is important in part because the products of the search including associations, somatic expression etc., are often not recognized by the client as being relevant (Ecker and Hulley, 1996). Recall the progressive “explicitization” of Karmiloff Smith requiring “scaffolding” as the knowledge is re-represented into a new structure. In this “bootstrapping” process the client doesn’t know the structure he is building, and therefore can’t always recognize the products of the search as being “building blocks” of the new structure. Therapist attention to capture and reflect back the “building blocks” is thus crucial.

As noted previously attentional resources for recall are allocated according to the goals of the person. Autobiographical memories are multifaceted constructs that contain conceptual as well as perceptual details. Recall of autobiographical events are thought to occur on a gradient of abstraction from conceptual to perceptually based episodic memory as a function of the goal of remembering (Sheldon et al., 2019). This results from specialization with the anterior hippocampus being activated in conceptual recall, while the posterior hippocampus is activated in perceptual recall. And it has been suggested that being able to recall along a gradient of abstraction depending on the goal is important in maximizing the adaptive function of memory. For example, a memory that is recalled conceptually without grounding in episodic detail has the potential to be deployed in a context inappropriate manner (van Kesteren and Meeter, 2020). This is often the case with the suboptimal priors identified in Coherence Therapy (Ecker, 2016). In contrast, the AIMCT argues that the product of Active Inference in Discovery is a (relatively) optimized schema that is capable of generalization but is grounded in context/episodic detail.

Successful Discovery in CT yields an explicit schema that drives symptom production and is typically characterized by three elements: an emotional wound, a presupposition, and protective actions (Ecker and Hulley, 1996). Emotional wounds reflect a continuum of stressful experiences up to and including overt psychological trauma. Presuppositions are unexamined assumptions taken from experience that are part of the subjective model of how the world is, or how it works. Protective actions serve to avoid any unwanted experience or event and can take an extremely wide range of forms e.g., dissociation, compulsion, low self-worth, depression, shame etc. Note that protective actions include mental actions as well as overt physical action. From the perspective of the AIMCT the explicit schema is the suboptimal prior that both illuminates the compromise of active inference in the client, and provides a path

to repair. To understand why, review of Active Inference under stress is helpful.

Active inference and stress

Faced with a threat, an individual must respond to ensure survival. This requires calculating expected free energy including epistemic and pragmatic terms (Linson et al., 2018). The individual must consider different policies or sequences of actions that will minimize uncertainty through an optimized combination of exploration and exploitation leading to the preferred outcome. Then the best one must be selected. He must try to choose the most viable path to his desired future.

Faced with a potential existential threat e.g., an active shooter, short term thermodynamic considerations are paramount (Linson et al., 2020). The imperative is to reduce expected free energy quickly [Hamilton's principle of least action (Parr et al., 2022)]. A policy to reduce uncertainty by exploration to better assess the threat runs the risk not only of putting oneself more directly in harm's way, but also in wasting valuable time and energy that could be used for escape. Individuals differ in how much evidence they seek before abandoning assessment and trying to escape. More formally, the difference in the precision of prior preferences for avoiding harm is a sensory evidence accumulation threshold that serves as a crossover point from exploration to exploitation. At the crossover point the individual stops considering "what is it" and acts on the prior belief of mortal danger to "get safe." Agent based modeling of this type of scenario has yielded some interesting results.

Agent based modeling of Active Inference under stress has been used to explore a model of PTSD (Linson et al., 2020). One important finding consistent with clinical phenomenology is that "a high negative preference (set by evolution or learning) for a stressor dependent outcome rapidly leads to exploitation (pragmatic foraging) before further state estimation." The resulting exploitation has been called a "stressor mitigation policy." "I really don't want to experience this bad thing so I don't bother to find out if I'm actually dealing with this bad thing, I just do this action to avoid the bad thing." And crucially, this may lead to a fearful cue being taken not as a possible state of affairs prompting attempts to confirm, but as confirmatory evidence itself. The curiosity reflected by foraging behavior has been replaced by a state of pragmatic defensive responding.

The greater the balance is tipped away from exploration, the greater the compromise of the inference process. In this way PTSD has been conceptualized as being a condition of stress induced impairment on a continuum, where prior beliefs of threat are not subject to updating by sensory samples (Linson and Friston, 2019). Inference Interrupted.

The AIMCT posits the following scenario as the pathogenesis of the typical Symptom Requiring Emotional

Truth retrieved in Coherence Therapy. In an "emotionally urgent situation in childhood," the client experiences extreme surprisal (unexpected uncertainty) with formidable computational complexity that precludes optimal inference. High Nor-epinephrine compromises frontal lobe function including evaluation of context and planning (Schacter et al., 2012; Arnsten et al., 2015; Fuster and Bressler, 2015). Unable to infer to the contrary, the threat may be treated as an existential threat. The complexity of calculating expected free energy for different courses of action makes policy selection challenging, and a policy is selected ("a protective action") to reduce uncertainty as rapidly as possible. (Free Energy reduction according to Hamilton's principle of least action). The policy selected is an *a priori* high probability "sure thing." In some cases, this may be an evolutionarily conserved response e.g., behavioral arrest or dissociation. A premium is placed on "never again" having that experience. If the "protective action" is successful, it persists as implicit procedural knowledge (Karmiloff-Smith Level I) to be used when needed. Computational complexity has been managed by a policy that requires minimal exploration and has minimal parameters e.g., no consideration of context. As a policy it is "good enough" (Gopnik et al., 2017) and is afforded high precision going forward because "it works" to avoid the dreaded situation.

The experience of extreme surprisal including physiological state, intractable complexity and negative affect is the "emotional wound" of the schema. Taken together these aversive elements prompt avoidance of the experience. This includes avoidance in thought. The individual avoids recall to be spared the interoception and pain of re-experiencing. Memory control blocks recall and the experience is unavailable for representational redescription. It remains implicit procedural knowledge in the system manifesting as automatic behavior when triggered by particular circumstance. Further Inference is precluded as the schema remains "modular" and isolated from other processors and types of knowledge. At that more benign end of the spectrum the schema may function as a suboptimal prior simply because it is never afforded the attention necessary for re-representation. "I never really thought it through."

As suggested by the SOMA framework strong schemas may operate automatically outside of conscious awareness and control as long as the effects on behavior are adaptive (Cleeremans et al., 2020). However, with attention the contents of a schema may be brought to awareness to gain control when behavior is not adaptive. The AIMCT posits that the implicit schema adopted in an emotionally urgent situation in childhood operates automatically and is adaptive, until it isn't. As the individual develops and matures her world becomes more complicated and circumstances change. From an objective function landscape perspective, the landscape has danced. What was once a local peak, a good enough solution, is now a valley. This necessitates search, and an

increase in the complexity in the modeling of her world in order to remain a good regulator of that world (Conant and Ross Ashby, 1970). When that doesn't occur due to blocked inference, unwanted consequences of her actions on the world begin to accumulate in the form of symptoms. Motivation to gain control builds. Coherence Therapy through an act of coordinated active inference repairs the inference process facilitating representational redescription with the emergence of context, agency and control.

Clinical case vignette

JS was a 62 y/o woman who presented with decades of horrific compulsive nightmares while dreaming, and while awake. These involved being physically tortured, critically injured in a variety of accidents, or sexually assaulted. This occurred several times a week while awake and while sleeping. For example, contemplating a drive to visit a friend she would imagine her car breaking down and being sexually assaulted, killed, and stuffed into the trunk by a man who stopped to help. Multiple courses of psychotherapy, medication, and behavioral interventions over the years had not significantly impacted these symptoms, despite ameliorating debilitating anxiety and depression. Coherence Therapy was initiated with the goal of "understanding how these symptoms might make sense." Discovery (active inference effecting representational re-description) yielded the following Emotional Truth (suboptimal prior). "My mother doesn't protect me, so it's critical that I imagine the worst case scenario so I am prepared for anything and can protect myself." The emotional wound was a suppressed memory of terror as she was driven in a car by her intoxicated father, with her mother's knowledge. The pre-supposition was "I have to do this by myself, no one else is going to help me." And "I need to consider anything that might ever happen." The protective action was to anticipate any and all possibilities that she could think of so she would have a plan for survival when she needed it, and would not be "caught off guard" (extreme surprisal) again.

Almost immediately following conscious expression the patient began to take exception with the presuppositions that contradicted her living knowledge. "My mother is deceased. I'm an adult and can protect myself. When my alcoholic ex-husband put my kids and I at risk I divorced him. I've been married for years to a loving husband who has helped me when I needed it. My life is good. I don't need to do this anymore."

Without further intervention the symptom frequency rapidly decreased over a period of weeks. The nightmares stopped. When the habitual thoughts did occur, they were rapidly followed by the thought "I don't need to do this anymore" and ceased. She remains symptom free at 4 year follow up.

From the existing conceptualization of CT the implicit emotional truth was "retrieved" from "memory systems other than those that hold one's explicit, autobiographical, episodic knowledge of past events" and made explicit (Ecker et al., 2012) (No physiological mechanisms or hypotheses offered to explain "retrieval," how schema becomes explicit etc.) The schema was presumed to have been "locked away by extraordinarily durable synapses." When the schema was confronted by contradictory living knowledge the mismatch triggered memory reconsolidation. This resulted in "unlearning" and "erasure" of the schema. "The process of unlearning and erasing that schema or mental model thoroughly resolves and puts to rest a core, personal theme of emotional distress..." (Ecker and Bridges, 2020). In summary, "the optimal process of psychotherapy consists of guiding the profound unlearning of the symptom-generating emotional learnings, nullifying and erasing them *via* the memory reconsolidation process" (Ecker and Bridges, 2020).

From the perspective of the AIMCT the client experiences a state of extreme surprisal with unexpected uncertainty (in a car being driven by drunk father). The computational complexity of calculating expected free energy to facilitate actions to ensure survival is formidable. Therefore a "stressor mitigation policy" is adopted to avoid ever finding herself in that situation again. This policy entails running simulations of an infinite set of "worst case scenarios." It is deployed without epistemic foraging to determine if it is necessary (Am I actually in danger at this time, in this context?). As a child with an alcoholic father and a mother with a demonstrated failure to protect her, it is a "good enough" policy that is adaptive.

The experience was terrifying and the policy adopted serves to avoid having the experience again. This includes re-experiencing the affect of terror and the physiological arousal accompanying recall of this episodic memory. Memory control mechanisms presumably involving the right dorsolateral PFC suppress recall and help to regulate affect in an (initially) adaptive action. The policy continues to run outside of awareness. The unintended consequences begin to accumulate including nightmares and compulsive simulation of personal danger. The active suppression of recall renders the episodic memory unavailable for active inference and representational redescription.

The accumulation of unintended consequences (symptoms) motivates seeking treatment. Guided by the framework of Bayes Optimal Pathology the therapist assigns high precision (confidence) to the existence of a suboptimal prior and initiates active inference to identify the hidden causes of the client's behavior. This is a goal directed "low confidence search" likely involving the frontal and dorsal parietal cortex. Motivation to eradicate the symptom prevails over motivation to suppress recall and the episodic memory is retrieved. "My father was drunk, it was terrifying." Through selective activation of sub-regions in the hippocampus driven by the specific purpose of remembering, the appropriate elements of memory (perceptual

vs. conceptual) are retrieved as the client's brain re-represents its own activity in an attempt to understand and gain control over the unintended consequences of its actions. The result is a model of herself with agency and control over the production of her symptoms.

From the initial event to the present the client's circumstances have changed significantly. The complexity of her situation has increased as she is now an adult with life experience, resources etc., that she didn't have as a child. Therefore her (old) model is no longer a good regulator of her environment. Rapidly recognizing the shift in context (no longer a dependent child) the problematic behavior is recognized as unnecessary (I don't need to do this anymore"). Rather than "unlearning" or "erasing" anything, she has learned a model of herself including motivation, behavior and consequences that contains an appreciation of its former utility and current irrelevance. This structure learning affords conscious access, agency, and control. And most importantly to the client, effortless cessation of the presenting symptom.

Conclusion

The AIMCT places Coherence Therapy in the framework of the Free Energy Principle and the Bayesian Brain through the lens of Bayes Optimal Pathology. Coordinated Active Inference identifies the putative Suboptimal Prior in the subject leading to the repair of inference driven by reduction in free energy through representational redescription that confers conscious access, agency, control, and symptom resolution. In the process Coherence Therapy recruits the motivation to access painful memories. A process theory was proposed to explain how shifts in memory control leading to de-repression might occur given empiric data regarding motivated control, search and memory in the brain. Potential questions for future research that follow include: is there evidence of a R dorsal frontal activity shift as motivation is recruited and a suppressed memory is retrieved?, Is the dorsal parietal cortex engaged in the putative "low confidence search" for the suboptimal prior?

The identification of the Suboptimal Prior in a subject may also be used as a step toward identification of phenotypes. Suboptimal priors between subjects often show common features e.g., "occurring early in life, in an emotionally charged situation" (Ecker, 2016). Such observations can in turn be used to model phenotypes that are then subjected to *in silico* experimentation (Schwartenbeck and Friston, 2016). The results of such experimentation may then inform improvements in clinical practice e.g., extreme emotional stress may result in excessive precision of the suboptimal prior such that inference cannot be repaired through Coherence Therapy

without modulation of the precision through neurotransmitter manipulation like dopamine blockade (Friston K. J. et al., 2012). In other words, some people may fail to benefit from this form of psychotherapy until a specific neurodynamic state is shifted pharmacologically. Thus, within the framework of Active Inference, the clinical practice of Coherence Therapy can help guide the research agenda for Computational Psychiatry, the results of which can inform better practice (Corlett and Fletcher, 2014). A virtuous cycle of Dual Aspect Monism where the emergent properties of mind and brain resulting from the underlying neural dynamics reflect and inform each other. Psychodynamic Neuroscience.

Data availability statement

The original contributions presented in this study are included in this article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Anderson, H. G. H. (1992). "The client is the expert," in *Social construction and the therapeutic process*, ed. S. G. K. McNamee (Newbury Park, CA: Sage Publications), 25–39.
- Anderson, M. C., Ochsner, K. N., Kuhl, B., Cooper, J., Robertson, E., Gabrieli, S. W., et al. (2004). Neural systems underlying the suppression of unwanted memories. *Science* 303:232. doi: 10.1126/science.1089504
- Arnsten, A. F. T., Raskind, M. A., Taylor, F. B., and Connor, D. F. (2015). The effects of stress exposure on prefrontal cortex: Translating basic research into successful treatments for post-traumatic stress disorder. *Neurobiol. Stress* 1, 89–99. doi: 10.1016/j.ynstr.2014.10.002
- Barbas, H., and Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cereb. Cortex* 7, 635–646. doi: 10.1093/cercor/7.7.635
- Benoit, R. G., Hulbert, J. C., Huddleston, E., and Anderson, M. C. (2014). Adaptive top-down suppression of hippocampal activity and the purging of intrusive memories from consciousness. *J. Cogn. Neurosci.* 27, 96–111. doi: 10.1162/jocn_a_00696
- Benoit, R., and Anderson, M. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron* 76, 450–460. doi: 10.1016/j.neuron.2012.07.025
- Botella, S., and Botella, C. (2004). *The work of psychic figurability: Mental states without representation*. Abingdon: Routledge. doi: 10.4324/9780203342206
- Bowers, J. S., and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414. doi: 10.1037/a0026450
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Cabeza, R., Ciaramelli, E., Olson, I. R., and Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nat. Rev. Neurosci.* 9, 613–625. doi: 10.1038/nrn2459
- Carhart-Harris, R. L., and Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
- Catarino, A., Küpper, C. S., Werner-Seidler, A., Dalgleish, T., and Anderson, M. C. (2015). Failing to forget: Inhibitory-control deficits compromise memory suppression in posttraumatic stress disorder. *Psychol. Sci.* 26, 604–616. doi: 10.1177/0956797615569889
- Chamberlin, D. E. (2019a). The network balance model of trauma and resolution-level I: Large-scale neural Networks. *J. EMDR Pract. Res.* 13, 124–142. doi: 10.1891/1933-3196.13.2.124
- Chamberlin, D. E. (2019b). The predictive processing model of EMDR. *Front. Psychol.* 10:2267. doi: 10.3389/fpsyg.2019.02267
- Ciaramelli, E., Burianov, H., Vallesi, A., Cabeza, R., and Moscovitch, M. (2020). Functional interplay between posterior parietal cortex and hippocampus during detection of memory targets and non-targets. *Front. Neurosci.* 14:563768. doi: 10.3389/fnins.2020.563768
- Cieri, F., and Esposito, R. (2019). Psychoanalysis and neuroscience: The bridge between mind and brain. *Front. Psychol.* 10:1983. doi: 10.3389/fpsyg.2019.01983
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190217013.001.0001
- Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenol. Cogn. Sci.* 17, 521–534. doi: 10.1007/s11097-017-9525-z
- Clark, A., and Karmiloff-Smith, A. (1993). The Cognizer's innards: A philosophical and psychological perspective on the development of thought. *Mind Lang.* 8, 487–519. doi: 10.1111/j.1468-0017.1993.tb00299.x
- Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.-R., Muñoz-Moldes, S., et al. (2020). Learning to be conscious. *Trends Cogn. Sci.* 24, 112–123. doi: 10.1016/j.tics.2019.11.011
- Conant, R. C., and Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97. doi: 10.1080/00207727008920220
- Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017
- Corlett, P. R., and Fletcher, P. C. (2014). Computational psychiatry: A Rosetta stone linking the brain to mental illness. *Lancet Psychiatry* 1, 399–402. doi: 10.1016/S2215-0366(14)70298-6
- Danker, J. F., and Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychol. Bull.* 136, 87–102. doi: 10.1037/a0017937
- Daselaar, S. M., Rice, H. J., Greenberg, D. L., Cabeza, R., LaBar, K. S., and Rubin, D. C. (2008). The spatiotemporal dynamics of autobiographical memory: Neural correlates of recall, emotional intensity, and reliving. *Cereb. Cortex* 18, 217–229. doi: 10.1093/cercor/bhm048
- Daunizeau, J., den Ouden, H. E. M., Pessiglione, M., Kiebel, S. J., Stephan, K. E., and Friston, K. J. (2010). Observing the observer (I): Meta-Bayesian models of learning and decision-making. *PLoS One* 5:e15554. doi: 10.1371/journal.pone.0015554
- Depue, B. E., Orr, J. M., Smolker, H. R., Naaz, F., and Banich, M. T. (2015). The organization of right prefrontal networks reveals common mechanisms of inhibitory regulation across cognitive, emotional, and motor processes. *Cereb. Cortex* 26, 1634–1646. doi: 10.1093/cercor/bhu324
- Doya, K., Ishii, S., Pouget, A., and Rao, R. P. N. (2006). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262042383.001.0001
- Ecker, B. (2015). Memory reconsolidation understood and misunderstood. *Int. J. Neuropsychother.* 3, 2–46. doi: 10.12744/ijnpt.2015.0002-0046
- Ecker, B. (2016). *Coherence therapy practice manual*. Oakland, CA: Coherence Psychology Institute LLC.
- Ecker, B. (2018a). Clinical translation of memory reconsolidation research: Therapeutic methodology for transformational change by erasing implicit emotional learnings driving symptom production. *Int. J. Neuropsychother.* 6, 1–92. doi: 10.12744/ijnpt.2018.0001-0092
- Ecker, B. (2018b). "Workshop-coherence therapy," in *Proceedings of the psychotherapy networker conference*, Washington, DC.
- Ecker, B., and Bridges, S. K. (2020). How the science of memory reconsolidation advances the effectiveness and unification of psychotherapy. *Clin. Soc. Work J.* 48, 287–300. doi: 10.1007/s10615-020-00754-z
- Ecker, B., and Hulley, L. (1996). *Depth-oriented brief therapy: How to be brief when you were trained to be dee-and vice versa*. San Francisco, CA: Jossey-Bass Publishers.
- Ecker, B., and Toomey, B. (2008). Depotentiation of symptom-producing implicit memory in coherence therapy. *J. Construct. Psychol.* 21, 87–150. doi: 10.1080/10720530701853685
- Ecker, B., Ticic, R., and Hulley, L. (2012). *Unlocking the emotional brain: Eliminating symptoms at their roots using memory reconsolidation*. New York, NY: Routledge. doi: 10.4324/9780203804377
- Eliasmith, C. (2003). Moving beyond metaphors: Understanding the mind for what it is. *J. Philos.* 100, 493–520. doi: 10.5840/jphil2003100102
- Engen, H. G., and Anderson, M. C. (2018). Memory control: A fundamental mechanism of emotion regulation. *Trends Cogn. Sci.* 22, 982–995. doi: 10.1016/j.tics.2018.07.015
- Engen, H. G., Kanske, P., and Singer, T. (2017). The neural component-process architecture of endogenously generated emotion. *Soc. Cogn. Affect. Neurosci.* 12, 197–211. doi: 10.1093/scan/nsw108
- FitzGerald, T. H., Dolan, R. J., and Friston, K. J. (2014). Model averaging, optimal inference, and habit formation. *Front. Hum. Neurosci.* 8:457. doi: 10.3389/fnhum.2014.00457
- Fleming, S. M. (2020). Awareness as inference in a higher-order state space. *Neurosci. Conscious* 2020:niz020. doi: 10.1093/nc/niz020
- Freud, S. (1895). *A project for a scientific psychology*, Standard Edn, Vol. 1. London: Hogarth Press, 283–397.
- Freud, S., Strachey, J., Freud, A., Rothgeb, C. L., Richards, A., Strachey, A., et al. (1961). *The standard edition of the complete psychological works of Sigmund Freud*. London: Hogarth Press.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005
- Friston, K. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: A process theory. *Neural Comput.* 29, 1–49. doi: 10.1162/NECO_a_00912
- Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., and Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Comput.* 29, 2633–2683. doi: 10.1162/neco_a_00999
- Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Price, C. J., and Holmes, E. (2020). Generative models, linguistic communication and active inference. *Neurosci. Biobehav. Rev.* 118, 42–64. doi: 10.1016/j.neubiorev.2020.07.005
- Friston, K., Adams, R. A., Perrinet, L., and Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Front. Psychol.* 3:151. doi: 10.3389/fpsyg.2012.00151
- Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012). Dopamine, affordance and active inference. *PLoS. Comput. Biol.* 8:e1002327. doi: 10.1371/journal.pcbi.1002327
- Friston, K., and Frith, C. (2015). A duet for one. *Conscious Cogn.* 36, 390–405. doi: 10.1016/j.concog.2014.12.003
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Doherty, J. O., and Pezzulo, G. (2016). Active inference and learning. *Neurosci. Biobehav. Rev.* 68, 862–879. doi: 10.1016/j.neubiorev.2016.06.022
- Frith, C. D. (2012). The role of metacognition in human social interactions. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 367, 2213–2223. doi: 10.1098/rstb.2012.0123
- Fuster, J. M., and Bressler, S. L. (2015). Past makes future: Role of pFC in prediction. *J. Cogn. Neurosci.* 27, 639–654. doi: 10.1162/jocn_a_00746
- Gagnepain, P., Hulbert, J., and Anderson, M. C. (2017). Parallel regulation of memory and emotion supports the suppression of intrusive memories. *J. Neurosci.* 37:6423. doi: 10.1523/JNEUROSCI.2732-16.2017
- Gopnik, A., and Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychol. Bull.* 138, 1085–1108. doi: 10.1037/a0028044
- Gopnik, A., Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., et al. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proc Natl. Acad. Sci. U.S.A.* 114, 7892–7899. doi: 10.1073/pnas.1700811114
- Hobson, J. A., Gott, J. A., and Friston, K. J. (2021). Minds and brains, sleep and psychiatry. *Psychiatr. Res. Clin. Pract.* 3, 12–28. doi: 10.1176/appi.prp.20200023
- Hohwy, J. (2013). *The predictive mind*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001
- Holmes, J., and Nolte, T. (2019). “Surprise” and the Bayesian brain: Implications for psychotherapy theory and practice. *Front. Psychol.* 10:592. doi: 10.3389/fpsyg.2019.00592
- Hopkins, J. (2012). “Psychoanalysis, representation and neuroscience: The Freudian unconscious and the Bayesian brain,” in *From the couch to the lab: Psychoanalysis, neuroscience and cognitive psychology in dialogue*, eds A. Fotopoulou, D. Pfaff, and M. Conway (Oxford: Oxford University Press), 230–265. doi: 10.1093/med/9780199600526.003.0014
- Hopkins, J. (2016). Free energy and virtual reality in neuroscience and psychoanalysis: A complexity theory of dreaming and mental disorder. *Front. Psychol.* 7:922. doi: 10.3389/fpsyg.2016.00922
- Hulbert, J. C., Henson, R. N., and Anderson, M. C. (2016). Inducing amnesia through systemic suppression. *Nat. Commun.* 7:11003. doi: 10.1038/ncomms11003
- Ingvar, D. H. (1985). “Memory of the future”: An essay on the temporal organization of conscious awareness. *Hum. Neurobiol.* 4, 127–136.
- Johnson, B., and Flores Mosri, D. (2016). The neuropsychanalytic approach: Using neuroscience as the basic science of psychoanalysis. *Front. Psychol.* 7:1459. doi: 10.3389/fpsyg.2016.01459
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.
- Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children’s metalinguistic and repair data. *Cognition* 23, 95–147. doi: 10.1016/0010-0277(86)90040-5
- Karmiloff-Smith, A. (1994). Précis of beyond modularity: A developmental perspective on cognitive science. *Behav. Brain Sci.* 17, 693–707. doi: 10.1017/S0140525X00036621
- Kelly, G. (1991). *The psychology of personal constructs*. London: Routledge.
- Kruglanski, A. W., Jasko, K., and Friston, K. (2020). All thinking is ‘Wishful’ thinking. *Trends Cogn. Sci.* 24, 413–424. doi: 10.1016/j.tics.2020.03.004
- Lanius, R. A., Frewen, P. A., Tursich, M., Jetly, R., and McKinnon, M. C. (2015). Restoring large-scale brain networks in PTSD and related disorders: A proposal for neuroscientifically-informed treatment interventions. *Eur. J. Psychotraumatol.* 6:27313. doi: 10.3402/ejpt.v6.27313
- Levine, H. B. (2012). The colourless canvas: Representation, therapeutic action and the creation of mind. *Int. J. Psychoanal.* 93, 607–647. doi: 10.1111/j.1745-8315.2012.00574.x
- Linson, A., and Friston, K. (2019). Reframing PTSD for computational psychiatry with the active inference framework. *Cogn. Neuropsychiatry* 24, 347–368. doi: 10.1080/13546805.2019.1665994
- Linson, A., Clark, A., Ramamoorthy, S., and Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Front. Robot. AI* 5:21. doi: 10.3389/frobt.2018.00021
- Linson, A., Parr, T., and Friston, K. J. (2020). Active inference, stressors, and psychological trauma: A neuroethological model of (mal)adaptive explore-exploit dynamics in ecological context. *Behav. Brain Res.* 380:112421. doi: 10.1016/j.bbr.2019.112421
- Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., et al. (2020). Resilience after trauma: The role of memory suppression. *Science* 367:eaay8477. doi: 10.1126/science.aay8477
- Murray, G. B. (1992). Limbic music. *Psychosom. J. Consult. Liaison Psychiatry* 33, 16–23. doi: 10.1016/S0033-3182(92)72016-7
- Panagiotaropoulos, T. I., Wang, L., and Dehaene, S. (2020). Hierarchical architecture of conscious processing and subjective experience. *Cogn. Neuropsychol.* 37, 180–183. doi: 10.1080/02643294.2020.1760811
- Parr, T., and Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia* 104, 92–101. doi: 10.1016/j.neuropsychologia.2017.08.003
- Parr, T., and Pezzulo, G. (2021). Understanding, explanation, and active inference. *Front. Syst. Neurosci.* 15:772641. doi: 10.3389/fnsys.2021.772641
- Parr, T., Pezzulo, G., and Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/12441.001.0001
- Parr, T., Rees, G., and Friston, K. J. (2018). Computational neuropsychology and bayesian inference. *Front. Hum. Neurosci.* 12:61. doi: 10.3389/fnhum.2018.00061
- Paz-Alonso, P. M., Bunge, S. A., Anderson, M. C., and Ghetti, S. (2013). Strength of Coupling within a mnemonic control network differentiates those who can and cannot suppress memory retrieval. *J. Neurosci.* 33, 5017–5026. doi: 10.1523/JNEUROSCI.3459-12.2013
- Pezzulo, G., Rigoli, F., and Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends Cogn. Sci.* 22, 294–306. doi: 10.1016/j.tics.2018.01.009
- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., and Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron* 76, 677–694. doi: 10.1016/j.neuron.2012.11.001
- Schacter, D. L., Addis, D. R., and Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nat. Rev. Neurosci.* 8, 657–661. doi: 10.1038/nrn2213
- Schwartenbeck, P., and Friston, K. (2016). Computational phenotyping in psychiatry: A worked example. *eNeuro* 3, 1–18. doi: 10.1523/ENEURO.0049-16.2016
- Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., et al. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Med. Hypotheses* 84, 109–117. doi: 10.1016/j.mehy.2014.12.007
- Sheldon, S., Fenerci, C., and Gurguryan, L. (2019). A neurocognitive perspective on the forms and functions of autobiographical memory retrieval. *Front. Syst. Neurosci.* 13:4. doi: 10.3389/fnsys.2019.00004
- Shimamura, A. P. (2011). Episodic retrieval and the cortical binding of relational activity. *Cogn. Affect. Behav. Neurosci.* 11, 277–291. doi: 10.3758/s13415-011-0031-4
- Singh, L., Schüpbach, L., Moser, D. A., Wiest, R., Hermans, E. J., and Aue, T. (2020). The effect of optimistic expectancies on attention bias: Neural and behavioral correlates. *Sci. Rep.* 10:6495. doi: 10.1038/s41598-020-61440-1
- Smith, R., Alkozei, A., Lane, R. D., and Killgore, W. D. S. (2016). Unwanted reminders: The effects of emotional memory suppression on subsequent neurocognitive processing. *Conscious Cogn.* 44, 103–113. doi: 10.1016/j.concog.2016.07.008
- Smith, R., Schwartenbeck, P., Parr, T., and Friston, K. J. (2020). An active inference approach to modeling structure learning: Concept learning as an example case. *Front. Comput. Neurosci.* 14:41. doi: 10.3389/fncom.2020.00041

- Solms, M., and Turnbull, O. H. (2011). What is neuropsychanalysis. *Neuropsychanalysis* 13, 133–145. doi: 10.1080/15294145.2011.10773670
- Streb, M., Mecklinger, A., Anderson, M. C., Johanna, L.-H., and Michael, T. (2016). Memory control ability modulates intrusive memories after analogue trauma. *J. Affect. Disord.* 192, 134–142. doi: 10.1016/j.jad.2015.12.032
- Strelnikov, K. (2014). Neuroenergetics at the brain-mind interface: A conceptual approach. *Cogn. Process* 15, 297–306. doi: 10.1007/s10339-014-0609-1
- Toomey, B., and Ecker, B. (2007). Of neurons and knowings: Constructivism, coherence psychology, and their neurodynamic substrates. *J. Construct. Psychol.* 20, 201–245. doi: 10.1080/10720530701347860
- Topolinski, S., Likowski, K. U., Weyers, P., and Strack, F. (2009). The face of fluency: Semantic coherence automatically elicits a specific pattern of facial muscle reactions. *Cogn. Emot.* 23, 260–271. doi: 10.1080/02699930801994112
- Tucker, D. M., and Luu, P. (2021). Motive control of unconscious inference: The limbic base of adaptive Bayes. *Neurosci. Biobehav. Rev.* 128, 328–345. doi: 10.1016/j.neubiorev.2021.05.029
- van Kesteren, M. T. R., and Meeter, M. (2020). How to optimize knowledge construction in the brain. *NPJ Sci. Learn.* 5:5. doi: 10.1038/s41539-020-0064-y
- Wald, A. (1947). An essentially complete class of admissible decision functions. *Ann. Math. Stat.* 10, 549–555. doi: 10.1214/aoms/1177730345
- Zhang, P., Jamison, K., Engel, S., He, B., and He, S. (2011). Binocular rivalry requires visual attention. *Neuron* 71, 362–369. doi: 10.1016/j.neuron.2011.05.035



OPEN ACCESS

EDITED BY
Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY
Eamonn Walsh,
King's College London,
United Kingdom
Tamara Fischmann,
International Psychoanalytic University
Berlin, Germany

*CORRESPONDENCE
Anna Buchheim
✉ anna.buchheim@uibk.ac.at

SPECIALTY SECTION
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 26 September 2022

ACCEPTED 02 December 2022

PUBLISHED 04 January 2023

CITATION

Buchheim A, Kernberg OF, Netzer N,
Buchheim P, Perchtold-Stefan C,
Sperner-Unterweger B,
Beckenbauer F and Labek K (2023)
Differential neural response
to psychoanalytic intervention
techniques during structural
interviewing: A single-case analysis
using EEG.
Front. Hum. Neurosci. 16:1054518.
doi: 10.3389/fnhum.2022.1054518

COPYRIGHT

© 2023 Buchheim, Kernberg, Netzer,
Buchheim, Perchtold-Stefan,
Sperner-Unterweger, Beckenbauer
and Labek. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Differential neural response to psychoanalytic intervention techniques during structural interviewing: A single-case analysis using EEG

Anna Buchheim^{1*}, Otto F. Kernberg², Nikolaus Netzer^{3,4},
Peter Buchheim⁵, Corinna Perchtold-Stefan⁶,
Barbara Sperner-Unterweger⁷, Fabian Beckenbauer¹ and
Karin Labek¹

¹Institute of Psychology, University of Innsbruck, Innsbruck, Austria, ²New York-Presbyterian Hospital–Westchester Division, New York, NY, United States, ³Hermann Buhl Institute, University of Innsbruck, Eurac Research, Innsbruck, Austria, ⁴Eurac Research, Bolzano, Italy, ⁵Department of Psychosomatic Medicine and Psychotherapy, Technical University of Munich, Munich, Germany, ⁶Institute of Psychology, University of Graz, Graz, Austria, ⁷Department of Psychiatry, Psychotherapy and Psychosomatics, Medical University Innsbruck, Innsbruck, Austria

Background: Previous studies detected changes in the electroencephalographic (EEG) signal as an effect of psychoanalytic interventions. However, no study has investigated neural correlates of specific psychoanalytic interventions in the EEG power spectrum yet. In the present case study, we contrasted three types of interventions (clarification, confrontation, and interpretation) and a neutral control condition during a structural psychoanalytic interview conducted while EEG was recorded.

Methods: A 27-year-old male patient diagnosed with major depressive disorder and borderline personality disorder with recurrent suicidal and self-injurious behavior underwent a structural interview while recording EEG. Two independent experts selected by consensus the characteristic episodes of the four conditions (clarification, confrontation, interpretation, and neutral control) within the interview, which were included in the EEG analyses. Fast Fourier transformation (FFT) was applied to subsegments of the intervention type to analyze the EEG power spectra. Alpha and beta power from central, frontal, and parietal sites were considered in linear mixed-effects models with segments as a random factor with maximum-likelihood estimates due to the lack of balance in the length of the interview segments.

Results: The interventions “interpretation” and “confrontation” showed a significantly lower alpha power compared with the control condition in the central electrodes. In the frontal and parietal sites of the alpha power and all beta power sites, the omnibus tests (full model/model without intervention) and comparisons relative to control conditions showed no significant overall result or failed significance after alpha error correction.

Conclusion: Incisive interventions, such as confrontation with discrepancies and interpretation of unconscious intrapsychic conflicts, may have provoked temporary emotional lability, leading to a change in psychic processing akin to interference from external stimuli. This conclusion is consistent with the finding that interpretations, which are potentially the most concise interventions, had the strongest effects on alpha power. Using EEG during therapeutic psychoanalytic intervention techniques might be a helpful tool to evaluate differential responses to the psychotherapeutic process on a neural level. However, this single-case result has to be replicated in a larger sample and does not allow generalizations.

KEYWORDS

structural interviewing, psychoanalytic intervention techniques, clarification, confrontation, interpretation, EEG analysis, alpha power

1. Introduction

1.1. Basic principles of psychoanalytic treatment

Psychoanalytic psychotherapies pursue change strategies related to unconscious conflicts, defensive processes, personality functioning, and structure, as well as typical maladaptive behavior patterns. They refer to the identification and processing of dysfunctional relational, experiential, and behavioral patterns of the therapeutic relationship and the processes of transference and countertransference between a therapist and a patient (Thomä and Kächele, 1985). From a psychoanalytic perspective, it is assumed that it is not primarily the elaboration of an alternative behavioral strategy that leads to personality change but rather the ability to reflect on and re-evaluate the underlying problems of behavior and of contradictory self-representations that creates the conditions for lasting change. The psychodynamic focus is either on more restricted abilities for self-regulation and relationship regulation or on the inability to recognize, understand, reflect, and mentalize one's inner processes and those of others (Kernberg, 1999; Bateman and Fonagy, 2004).

Hereby, four aspects jointly constitute the very essence of psychoanalytic techniques: Interpretation, transference analysis, technical neutrality, and countertransference analysis (Etchegoyen, 1999; Kernberg, 2016). Interpretation is defined as a verbal communication by an analyst describing the hypothesis of an unconscious conflict that seems to have dominantly emerged in the patient's communication in the therapeutic encounter. The interpretative process may be classified into clarification, confrontation, and interpretation. Clarifications and general questions rely on the material more explicitly mentioned in the patient's narratives (conscious or preconscious), which is associated with conscious processes.

By contrast, the defining characteristic of confrontation and interpretation refers to the unconscious or preconscious material. Confrontations pave the way for analytic work, as they draw attention to important topics, indicate potential similarities (or inconsistencies) in patient narratives, and test the ability of the patient to initiate a process of self-reflection. Interpretations offer possible hypotheses beyond what the patient already knows and may unravel a patient's conflict or uncover motives of behavior. This condensed hypothesis is that interpretation "in the here and now," to be followed or completed with interpretation "in the there and then," that is, the genetic aspects of interpretation that refer to the patient's past linking the unconscious aspects of the present with the unconscious aspects of the past. Transference may be defined as the unconscious repetition in the here and now of pathogenic conflicts from the past, and the analysis of transference is the main source of specific change brought about by psychoanalytic treatment (Kernberg, 2016).

There is a range of manualized psychoanalytic therapies that vary in the extent to which they focus on more supportive or interpretive elements (e.g., transference-focused psychotherapy by Kernberg et al. (1989), Clarkin et al. (1999), panic-focused psychodynamic psychotherapy by Milrod et al. (1997), Mentalization Based Treatment by Bateman and Fonagy (2016). These manualized approaches share the core meaning of a comprehensive initial diagnostic assessment for treatment planning. Patients with more severe disturbances in personality functioning require more highly structured treatment settings, and frequently, the establishment of a formal treatment contract is necessary to minimize destructive attacks against self, others, and the treatment. Clinical interviews focusing on personality functioning, including several domains such as self and interpersonal functioning as the central tenet of personality pathology, provide a fruitful diagnostic approach. This is shared

by many theoretical orientations and evidence-based treatments (Pincus et al., 2020).

1.2. The structural interview as a diagnostic tool to assess personality structure and functioning

In structural interviewing, Kernberg (1981, p 169) explained: “structural diagnosis as an overall intrapsychic organization that provides stability, continuity throughout time and is facilitated by a special type of clinical interview—a ‘structural interview’—that focuses sharply on the relation between the interaction of patient and interviewer, the patient’s interpersonal functioning in general and the history of the present illness.” In severe personality disorders, Kernberg (1984) focused on the structural characteristics of Borderline Personality Organization. In psychodynamic therapy for personality pathology (Caligor et al., 2018), the structural interview is divided into the following: Phase I: Presenting complaints and symptoms, phase II: Personality functioning, phase III: Identity formation, phase IV: Past history, and final phase: Outstanding issues and questions. The structural interview stands for a pioneering, integrative approach in which, besides psychopathology and psychodynamics, personality structure and functioning on a continuum of severity of pathology on different levels of personality organization can be assessed (Buchheim et al., 1987). It combines psychopathology with a psychoanalytic focus on the patient–interviewer interaction. In the initial phase, the interviewer presents four questions in sequence about the main complaints and problems, their development, expectations of treatment, and current state of health and focuses on the clarification of the symptoms, the pathological character traits, and the conflicts or difficulties offered by the patient. In the middle phase, the interviewer asks the patient to characterize himself and describe the important reference persons, to get a picture of the internalized self and object relations, to understand the extent of identity diffusion or the capacity for identity integration, and to understand the typical defense mechanisms. The aim of the interview was also to gain a comprehensive picture of the patient’s essential areas of life, such as work, leisure, family, partnership, and sexuality. In addition, the ability to do a reality check to have a distinction between psychotic symptoms and disorders is always assessed.

By focusing on the interaction between the patient and the interviewer with the help of the psychoanalytic techniques of *clarification*, *confrontation*, and *interpretation* of conflictual themes and defense mechanisms, it is possible to work out the symptomatology and the underlying personality structure at the same time. A salient feature of the structural interview is a circular approach to questioning. Repeatedly returning to previously expressed but not yet sufficiently clarified symptoms and problems allows the interviewer to return to the starting point and reinitiate a new circle of inquiry, evaluating

contradictions in different contexts at a later time and forming a more complete picture.

The structural diagnosis depends upon how the patient reacts and handles clarifications, confrontations, and interpretations in the interview. This interview technique is a challenge for both interviewer and patient and can lead to a kind of “experimental labilization,” through which the lowered functional level of the personality organization (“bottom of the rock”) under stressful conditions can also become recognizable in the interview.

The focus on the severity of self- and interpersonal dysfunction represents a considerable step forward in making diagnoses clinically meaningful (Blüml and Doering, 2021). With the alternative model, the DSM-5 recognizes the need for a dimensional approach to personality pathology and functioning, which posits self and interpersonal functioning as key defining dimensions of both normal personality and PDs, along a continuum of severity of dysfunction in these domains (Hörz-Sagstetter et al., 2018). This new development underlines the lasting topicality of Kernberg’s theoretical approach.

1.3. Research on psychoanalytic constructs and effects of treatment

There is strong, accumulating evidence from attachment theory, experimental psychology research, including neuroscience, and psychotherapy research that support the validity and clinical usefulness of several psychoanalytic basic constructs (e.g., defensive processes, transference and countertransference, insight, and mentalizing) that underlie psychodynamic psychotherapy (Yakeley, 2018; Levy et al., 2019).

Recently, three trials of psychodynamic therapies were structured to focus predominantly on the therapeutic relationship and transference. The first experimental study of transference compared psychodynamic therapy by randomizing patients to receive or not receive transference interpretations (Høglend et al., 2008). In the subsample of patients with poor personality functioning (predominantly cluster C personality disorders), therapy without transference interpretations was less effective at improving patients’ insights into their conflicts and patterns of defense, leading to worse improvements in psychosocial functioning compared to therapy with consistent use of transference interpretations (Høglend et al., 2008, 2011). Consistent with the integrative view for individuals with lower levels of personality functioning, transference interpretations were especially helpful if the therapists acted from a “parental” stance, whereas the contrary was true for individuals with higher levels of personality functioning. Among samples of patients with borderline personality disorder (BPD), transference-focused psychotherapy (TFP) has been compared to dialectical-behavioral therapy, dynamic-supportive therapy, and treatment by community experts (Clarkin et al., 2007; Doering et al., 2010). In both trials, TFP was also shown to

uniquely promote patient improvements in mentalization and attachment security (Levy et al., 2006; Fischer-Kern et al., 2015; Buchheim et al., 2017), which is consistent with a view that transference interpretation may be uniquely helpful in this population specifically for fostering intrapsychic integration.

Recent developments in neuroscience have fertilized and intensified an interdisciplinary dialogue between psychoanalysis and neuroscience (e.g., Kandel, 1998, 2013; Carhart-Harris and Friston, 2010; Prosser et al., 2018; Solms, 2021a,b). The cooperation between the two disciplines has resulted in numerous experimental studies that shed new light on psychoanalytic constructs and techniques (e.g., Solms, 2011; Panksepp and Solms, 2012; Böker et al., 2013; Shevrin et al., 2013). One study group recently focused on neural responses on free association in healthy individuals (Kehyayan et al., 2013; Schmeing et al., 2013). Moreover, several fMRI studies demonstrated the effects of psychodynamic treatment on a neural level (Beutel et al., 2010, 2012; Buchheim et al., 2012; de Greck et al., 2011, 2013; see review Abbass et al., 2014; Messina et al., 2016; Perez et al., 2016). Only a few studies examined neural changes during psychodynamic treatment using electroencephalographic (EEG) (Unterrainer et al., 2013, 2014; Buchheim et al., 2018). Buchheim et al. (2018) demonstrated that, at the beginning of the treatment, patients showed significantly higher late positive potentials (LPPs) at the frontocentral sites and sustained gamma-band activity compared to the controls. After 15 months of treatment, LPP amplitudes and gamma-band responses of the patients decreased and equalized to the amplitudes of the healthy controls. Here, LPP and gamma-band activity were considered potential endophenotypes of the processing of emotional content in the course of psychoanalytic treatment.

However, no study so far has investigated the neural responses to basic psychoanalytic interventions (clarification, confrontation, and interpretation) in a standardized interview setting. For the study of awake humans, alpha, beta, and gamma waves were central frequency bands observed in various tasks. For example, alpha rhythms (8–12 Hz) are highly responsive to sensory stimuli and motor tasks (Klimesch et al., 1996; Williamson et al., 1997; Cohen, 2017). Beta oscillations (13–30 Hz) have been identified in many perceptual, cognitive, and motor processes in various EEG studies (e.g., Schmidt et al., 2019). Gamma-band waves were excluded from further analyses to carefully account for possible confounds with muscle artifacts occurring in a similar frequency range. The aim of the present study was to analyze for the first time a patient's brain activity during a structural interview by contrasting clarification, confrontation, and interpretation techniques using EEG.

2. Materials and methods

The patient and the therapist gave written informed consent to the analysis and publication of the data. However, the case

report was anonymized to protect the patient's identity. The case study was approved by the ethical review board of the University Innsbruck. The structural interview was administered by Otto Kernberg, the senior expert in the interviewing procedure, who had no information about the patient.

2.1. Participant

The 27-year-old male patient was diagnosed with major depression with past suicidal attempts and was in inpatient treatment. The patient has lost his father, who committed suicide 6 years ago. The patient associated the beginning of his symptoms (depressive mood, self-mutilating behavior, suicidal thoughts, and feelings of emptiness) with severe loss. Since then, the patient has lost all his interests, broke off his studies at university, and had relationship problems with significant others. His temporary promiscuous behavior especially endangered his relationship with his girlfriend. Before the start of the interview procedure, the patient and the therapist agreed to participate in the study. Neither the therapist nor the patient had any information about each other and did not receive payment for conducting or participating the interview. The interview was assessed in the initial phase of the inpatient treatment. Both were given general information about the study and signed a declaration of their willingness to participate in the interview and the EEG recordings.

2.2. Interview procedure

The interview was conducted in the standardized format described by Kernberg (1981). After a multi-part initial questionnaire about the main complaints and problems, their development, expectations of treatment, and current state of health, the interviewer, Kernberg, initially focused on the symptoms, conflicts, or difficulties offered by the patient. The interviewer *clarified* to understand the main symptoms like depressive mood, strong self-injury, and concentration problems. He continued to *clarify* if there were any other problems besides the depressive symptoms and self-harm. As a result, the patient reported the suicide of his father 6 years ago with no feelings of mourning but of being shocked. The next clarification process focused on the concentration problems in his studies and his failure to succeed in examinations. Along with the circular approach to questioning, the interviewer again *clarified* if there were any other problems besides the reported ones. The patient first denied having any other problems several times but then reported the negative effects of his self-harm behavior on his long-lasting relationship with his girlfriend. In this context, he finally confessed that he had cheated on her with other women, which had led to an ongoing severe crisis by the time of the interview. In the further course of the interview,

the interviewer asked the patient to characterize himself and his girlfriend to get a picture of the internalized self and object relations and the extent of identity diffusion. Here, the first contradictions occurred: On the one side, the girlfriend was described as available and empathic, with a relationship in which both could share a lot, including satisfying sexuality. On the other side, the patient reported about the senselessness in his life with no support from anyone, including his parents. The interviewer now *confronted* him several times that he denied the existence of his girlfriend and that he prefers to harm himself instead of studying. He confronted the patient with his way of mourning by being shocked but without any feelings of grief. The patient then realized that he felt abandoned by his father. The interviewer again confronted the patient that he was going to destroy his relationship by cheating on his girlfriend and that he was in danger of being abandoned. The patient could see this point and showed adequate feelings of guilt. In the next step, the interviewer summarized the core conflicts of the patient by confronting him again and giving several interpretations. He interpreted that the patient unconsciously did not dare to have a better life than his father and therefore is at risk of destroying himself and his relationships, especially the one with his girlfriend. He interpreted his lack of concern for himself and his inability to make any choices for change as a suboptimal compromise to avoid competing with his father, who was leading a most unhappy life. The patient was emotionally affected by these numerous interpretations and showed signs of insight and an increasing wish to change some major aspects of his life.

This interview technique led to the “experimental labilization” (Kernberg, 1981) described earlier, through which the lowered functional level of personality organization (borderline personality organization) with clear signs of identity diffusion, a disrupted concept of self, lack of care, concern, and responsibility for himself and others became evident under these activating stressful conditions in the interview.

2.3. EEG procedure

Before the interview, EEG electrodes were placed on the patient's head. The patient's EEG signal was measured continuously during the structural interview. During the interview, the patient and therapist sat facing each other. To achieve an exact match with the EEG signal, an audiovisual recording of the EEG and the interview was made simultaneously. The brain activity was recorded throughout the whole duration of the interview. Thus, the entire EEG recording lasted 63 min 35.7 s. To reduce possible artifacts, we used only the patient's EEG signal when the patient was not talking but paying attention to the therapist's treatment. Recording, preprocessing, and data analysis were performed using Brain Vision Recorder software (2.0, Brain Products,

Gilching, Germany) and Brain Vision Analyzer software (2.0, Brain Products, Gilching, Germany).

2.4. Expert ratings of the therapeutic interventions

To evaluate the therapist's interventions during the clinical interview, the entire interview was transcribed in the first step. The transcript was then used to assess the type of intervention (clarification, confrontation, interpretation, and the control condition) by two independent and experienced experts in the field. A joint consensus then finalized the results of the expert's assessments. The final and definitive assessments of the interventions were then included in the EEG analyses. Overall, the interview resulted in 28 interview sequences for confrontation, 91 for clarification, seven for interpretation, and 27 for the control condition. This resulted in 218 EEG segments for confrontation, 434 for clarification, 71 for interpretation, and 79 for control segments. For the control condition, neutral sentences or statements without emotional content were selected, such as “What does middle school mean?” or “My name is Prof. Kernberg, I am a psychiatrist from New York.” These segments have been used as a basis for the EEG analyses.

2.5. Acquisition and processing of EEG data

Electroencephalographic (EEG) activity was recorded at 30 electrode sites of the extended 10–20 system with vertical and horizontal electrooculogram (EOG) components. A set of 11 silver electrodes attached with a glue paste (Nihon Kohden Elefix EEG paste) to the scalp was used to record EEG signals. The impedances of the EEG electrodes were below 5 k Ω . EEG data were sampled at 2,500 Hz and subsequently down-sampled to 2,048 Hz. EEG data were carefully checked for artifact-contaminated signals (eye blinks, horizontal and vertical eye movements, muscle artifacts, etc.) by visual inspection. For further processing of the data, an average offline reference was computed. The data were filtered by excluding fluctuations below 1 Hz (time constant 3.0 s, 48 dB/oct) and above 45 Hz (48 dB/oct).

Based on the therapist's interventions, we divided the patient's artifact-free epochs of the EEG signal into “segments” according to intervention type: Clarification, confrontation, interpretation, and a control condition. These segments were further divided into non-overlapping subsegments of duration 1 s for spectral analysis. For artifact-free epochs, EEG power spectra (within a frequency range of 1–45 Hz) were computed using a fast Fourier transformation (FFT) on the subsegments of 1 s with a maximum resolution of 1 Hz after applying a 10% Hanning window. The signal sequence in each channel of

the EEG data was decomposed into four specific spectral bands (delta: 0.5–3.5 Hz, theta: 4–7.5 Hz, alpha: 8–12.5 Hz, and beta: 13–30 Hz). Only the alpha and beta bands were considered in the statistical analysis.

2.6. Statistical analysis

Logarithms of alpha and beta power μV^2 in frontal (F: Fp1 and Fp2), central (C: C3 and C4), and parieto-temporal (PT: P3 and P4) electrodes were taken and modeled as a function of intervention type (clarification, confrontation, interpretation, and the control condition) and laterality in a linear mixed-effects model with segments as a random factor. Since the interview segments differed in length, we relied on maximum-likelihood estimates of random effects to account for the lack of balance in the data (Gelman and Hill, 2006; Twisk, 2006). We adopted a Bonferroni-corrected significance threshold for the three electrode sites in omnibus χ^2 -tests of the effect of intervention type ($p < 0.017$). When the null could be rejected in the omnibus test, we followed up with planned comparisons of the intervention types relative to the control condition (two-tailed). The analysis was conducted using the freely available package R (The R Foundation for Statistical Computing,¹ Vienna, Austria; repeated measures regression: function lme, package nlme, Pinheiro and Bates (2000), mixed-effects models in S and S-PLUS. Berlin: Springer.). Boxplots were drawn with the freely available package ggplot2 (Wickham, 2016, Berlin, Springer).

3. Results

3.1. Clinical evaluation of the interview by the therapist

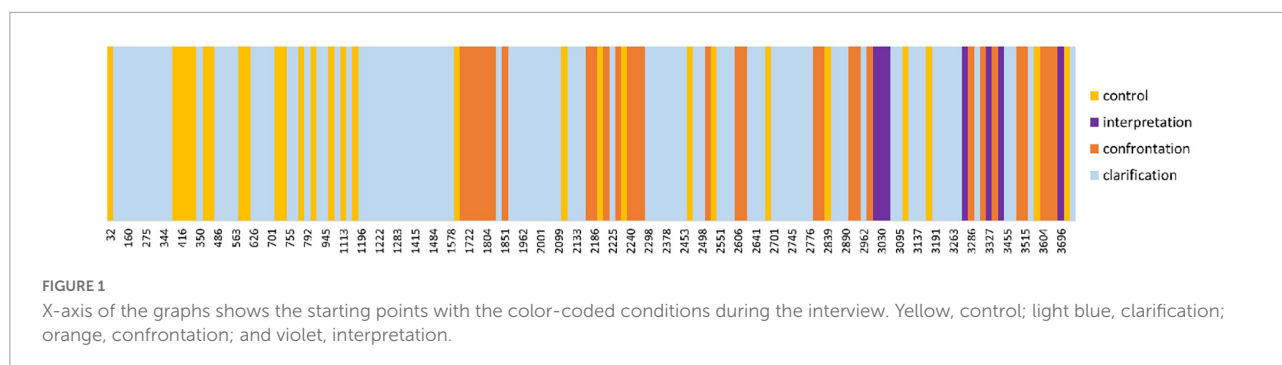
The 27-year-old male patient was diagnosed with major depression and depressive personality disorder. The structural

interview came to the structural diagnosis of a “borderline personality organization.” At this level, reality testing is intact. However, patients with a borderline level of personality organization have a fragmented sense of self and others (unlike the less severe neurotic organization with an integrated self). Because they possess a fragmented sense of self, they do not have a consistent view of themselves or others over time and across situations. This fragmented sense of self is the most significant and defining feature of the borderline level and results in severe and repetitive problems with interpersonal relationships. In our case, the patient has lost his father, who committed suicide 6 years ago. The patient’s fragmented sense of self and others became evident in that he could not mourn at all but was shocked at the same time. Moreover, the patient associated the beginning of his symptoms (depressive mood, self-mutilating behavior, suicidal thoughts, and feelings of emptiness) with this loss. Since then, the patient shown a lack of engagement at the university, a loss of interest, and developed relationship problems. In this context, he described an almost perfect relationship with his girlfriend but, at the same time, behaved in a completely contrary manner. He endangered the relationship with his temporary promiscuous behavior, which he denied and repressed. From a psychodynamic perspective, the patient demonstrated infantile tendencies, including clear signs of identity diffusion, a disrupted concept of self and others, and a lack of care, concern, and responsibility for oneself. While *clarification* episodes during the interview prepared for identifying contradictions and conflicts, *confrontation* and *interpretation* episodes aimed to reveal splitting and denial mechanisms, uncovering that the patient’s repressed guilt feeling toward his father may hinder him from living a fulfilling life.

3.2. EEG results

Based on the therapist’s interventions, we classified the patient’s artifact-free epochs of the EEG signal into “segments” according to intervention type: clarification, confrontation, interpretation, and a control condition. To obtain a precise estimate of the duration and distribution of the seconds/condition of the interview, we considered the total

¹ <http://www.r-project.org/>



duration of the interview and for each condition in the first step and calculated the talking and listening to the therapist durations of the patient in the second step. The total duration of the clarification condition was 622 s and started at second 39. The confrontation condition lasted 293 s in total and started with a first interaction at minute 28 and 5 s. The interpretation segments started at minute 50 and 11 s with a total time of 102 s, and the control condition lasted 122 s, starting at second 33 at the beginning of the interview (see **Figure 1**).

In the clarification condition, the patient spoke for 560 s (47.37%) and listened to the interviewer for 622 s (52.62%). In the confrontation condition, the evaluation of speaking and listening durations showed that the patient spoke for 136 s (31.70%) and listened to the interviewer for 293 s (68.30%). In the interpretation condition, the patient spoke for 64 s (39.29%) and listened to the interviewer for 102 s (61.91%). In the control condition, both speaking and listening durations lasted 122 s.

Based on the sentences identified by the expert ratings, a total of 1139 s was determined for EEG analysis. Of these, 340 s were excluded due to artifacts, resulting in a maximum of 799 s available for the statistical analysis, which were, in turn, divided between the different conditions. This resulted in a final duration time for each of the conditions: The clarification condition comprised a total of 622 s. Of these, 434 s were again included in the statistical analysis, for which 188 s were excluded as artifacts. The confrontation condition involved 293 s, of which 218 s were included in the statistical analysis and 75 s were omitted due to artifacts. The interpretation condition comprised a total of 102 s, of which 71 s were used in the statistical analysis and 31 s were excluded as artifacts. The control condition comprised 122 s, with 46 s omitted due to artifacts and 76 s used for statistics. A quasibinomial logistic regression revealed no significant result between the intervention interpretation [$\beta = -0.25$, $SE = 0.58$, $t(3,158) = -0.43$, n.s.], confrontation [$\beta = -0.45$, $SE = 0.36$, $t(3,158) = -1.24$, n.s.], clarification [$\beta = -0.28$, $SE = 0.29$, $t(3,158) = -0.99$, n.s.], and the control condition (see **Table 1**).

Participant's electrophysiological responses in the EEG power spectrum were analyzed in the alpha and beta power in the frontal, central, and parietal electrodes in the different interventions (condition: Clarification, confrontation, interpretation, and control condition) and the brain lateralization (left/right hemisphere) (see **Table 2**).

3.2.1. Alpha power spectrum

An omnibus test for the effect of intervention type in the *central electrodes* revealed an overall significant result in the alpha power spectrum [$\chi^2(3) = 11.95$, $p = 0.007$]. The intervention "interpretation" showed a significantly lower alpha power in comparison with the control condition [$\beta = -0.41$,

TABLE 1 Mean values, standard deviation (SD), min/max values, and subsegment counts (count seg.) of logarithms of alpha and beta power μV^2 in the frontal (Fp1; Fp2), central (C3 and C4), and parieto-temporal (P3 and P4) electrodes (mean values averaged across subsegments for each condition).

	Mean (SD)	Min
Alpha power		
Central electrodes		
Control	1.78 (0.68)	-0.06
Clarification	1.64 (0.60)	-0.83
Confrontation	1.47 (0.62)	-0.70
Interpretation	1.33 (0.67)	-0.29
Frontal electrodes		
Control	2.37 (0.73)	0.41
Clarification	2.36 (0.76)	0.23
Confrontation	2.15 (0.70)	0.52
Interpretation	2.06 (0.77)	0.20
Parieto-temporal electrodes		
Control	2.32 (0.65)	0.56
Clarification	2.28 (0.67)	-0.18
Confrontation	2.18 (0.69)	-0.21
Interpretation	2.02 (0.65)	0.81
Beta power		
Central electrodes		
Control	0.59 (0.30)	-0.25
Clarification	0.51 (0.35)	-0.55
Confrontation	0.54 (0.34)	-0.41
Interpretation	0.61 (0.33)	-0.20
Frontal electrodes		
Control	1.37 (0.63)	-0.07
Clarification	1.32 (0.63)	-0.26
Confrontation	1.19 (0.61)	-0.60
Interpretation	1.17 (0.84)	-0.88
Parieto-temporal electrodes		
Control	0.90 (0.37)	-0.12
Clarification	0.70 (0.44)	-0.31
Confrontation	0.76 (0.46)	-0.57
Interpretation	0.68 (0.37)	-0.22

$SE = 0.15$, $t(147) = -2.80$, $p = 0.006$] similar to the intervention "confrontation" [$\beta = -0.28$, $SE = 0.11$, $t(147) = -2.59$, $p = 0.01$]. By contrast, "clarification" showed no significant power difference to the control condition in the central electrodes [$\beta = -0.11$, $SE = 0.09$, $t(147) = -1.14$, n.s.]. Results of the logarithms of alpha power in the central electrodes of the interventions clarification, confrontation, interventions and

TABLE 2 Report of the interview length in seconds for the interventions assessed by the experts.

	Total sec	Analyzed sec	Removed sec	% Included
All conditions	1139	799	340	70.15
Clarification	622	434	188	69.77
Confrontation	293	218	75	74.40
Interpretation	102	71	31	69.61
Control	122	76	46	62.30

The first column shows the seconds in total for the electroencephalographic (EEG) analysis (total sec), the second column shows the analyzed seconds after artifact rejection (analyzed sec), and the third column the rejected parts in seconds (removed sec). The last column shows the percentage of the included interview parts in seconds after artifact rejection (% included).

control condition can be seen in **Figure 2A**. Alpha power was higher in the left than in the right hemisphere in these electrodes [$\beta = -0.37$, $SE = 0.04$, $t(1452) = -10.60$, $p < 0.001$]. There was no interaction with the intervention type (see **Figure 2**).

To account for changes in alpha power that occurred for the interview, a model adjusted for time was calculated. After adjusting for time [$\beta = -0.11$, $SE = 0.036$, $t(148) = -3.0$, $p = 0.0027$], the “interpretation” intervention [$\beta = -0.44$, $SE = 0.14$, $t(148) = -3.15$, $p = 0.002$], and confrontation [$\beta = -0.23$, $SE = 0.10$, $t(148) = -2.24$, $p = 0.026$] scored significantly lower in alpha power compared to the control condition. The “Clarification” intervention [$\beta = -0.06$, $SE = 0.10$, $t(148) = 0.52$, n.s.] still showed no significant difference compared with the control condition in the central electrodes. Similar results were obtained without the adjustment.

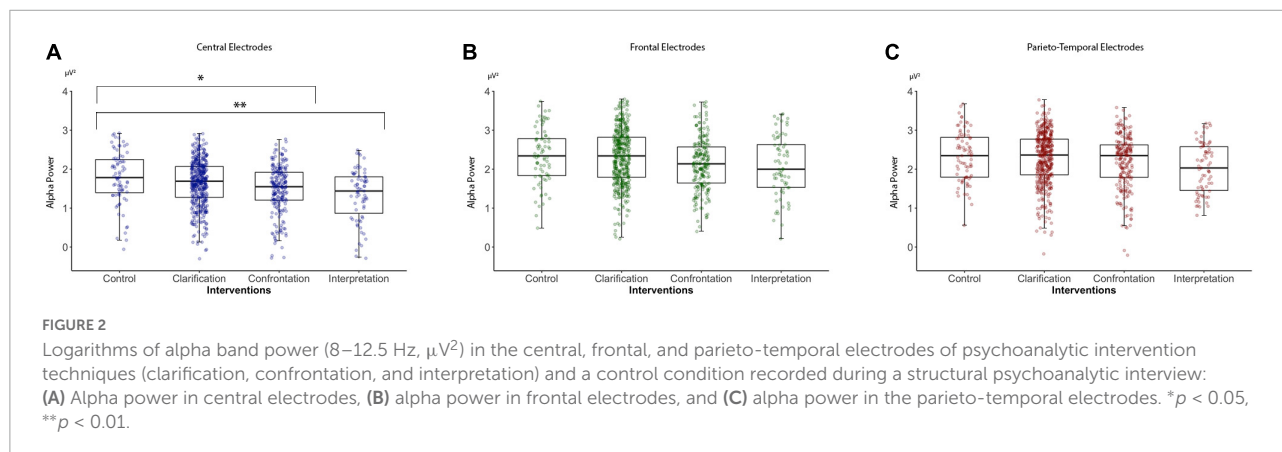
The analysis of the alpha power spectrum in the *frontal electrodes* revealed a trend for an effect of intervention type in the omnibus test [$\chi^2(3) = 9.58$, $p = 0.023$]. While failing to reach stringent significance thresholds, the effects of intervention type relative to the control condition were similar to those of the central electrodes. Alpha power was lower in

the condition “interpretation” than in the control condition [$\beta = -0.25$, $SE = 0.14$; $t(147) = -1.71$, $p = 0.089$] and during “confrontation” [$\beta = -0.14$, $SE = 0.10$; $t(147) = -1.31$, n.s.] but was no different than control during “clarification” [$\beta = 0.05$, $SE = 0.09$; $t(147) = 0.53$, n.s.]. Results of the logarithms of alpha power in the frontal electrodes of the interventions clarification, confrontation, interventions and control condition can be seen in **Figure 2B**. No hemispheric lateralization effect was found in the frontal electrodes [$\beta = -0.04$, $SE = 0.04$; $t(1451) = -1.23$, n.s.].

The analysis of the alpha power spectrum in the *parietal electrodes* failed to detect any significant effect of intervention type [$\chi^2(3) = 3.43$, n.s.] with no significant results in the intervention type interpretation [$\beta = -0.23$, $SE = 0.16$; $t(147) = -1.41$, n.s], confrontation [$\beta = -0.03$, $SE = 0.11$; $t(147) = -0.24$, n.s], and clarification [$\beta = 0.03$, $SE = 0.1$; $t(147) = 0.75$, n.s.]. Results of the logarithms of alpha power in the parietal electrodes of the interventions clarification, confrontation, interventions and control condition can be seen in **Figure 2C**. However, there was a significant lateralization effect [$\beta = -0.13$, $SE = 0.03$; $t(1452) = -3.90$, $p < 0.001$]. There was no interaction with the intervention type.

3.2.2. Beta power spectrum

Omnibus tests comparing a full model and a model without an intervention condition in the beta power gave no significant overall results in the central [$\chi^2(3) = 2.27$, n.s.], frontal [$\chi^2(3) = 5.85$, n.s], or parietal electrodes [$\chi^2(3) = 5.50$, n.s]. However, the planned comparisons relative to the control condition in the parietal electrodes exhibited an effect toward a lower beta spectral power in the condition “interpretation” [$\beta = -0.22$, $SE = 0.10$, $t(147) = -2.11$, $p < 0.037$], the interventions confrontation [$\beta = -0.14$, $SE = 0.07$, $t(147) = -1.87$, $p < 0.064$], and “clarification” [$\beta = -0.11$, $SE = 0.06$, $t(147) = -1.75$, $p < 0.082$]. Statistical significance failed after correction for multiple comparisons (see **Figure 2**).



4. Discussion

In this single-case study, we observed a decrease of power in the alpha band of a 32-electrode EEG during a psychoanalytic interview based on the technique of Kernberg (1981). This structured interview technique led to an emotional response, through which the lowered functional level of personality organization (borderline personality organization) with clear signs of identity diffusion, a disrupted concept of self, and a lack of care, concern, and responsibility for himself and others became evident in the interview. Interventional parts like confrontation and interpretation in a structured psychoanalytic interview and asking the patient to process personal unconscious conflicts have led to a destabilizing state and an impaired neurocognitive control mirrored by a loss in EEG alpha power.

In previous studies, decrements of alpha power in the EEG of awake participants were observed during the execution of cognitive tasks when subjected to the interference of external visual or acoustic signals or intervening limb movements. The visual signal is an externally evoked event that redirects EEG signals to posterior neuronal networks and can increase or decrease alpha power. For example, Pfurtscheller and Lopes da Silva (1999) described the basic principles of event-related EEG/MEG synchronization and desynchronization and that the other external motor sensor stimuli could have the same effect as visual stimuli. This result was confirmed by recent research by Hager et al. (2018). The decrements in alpha power of the present study suggest that the destructuring effect of confrontations and interpretations on this patient had an interfering effect on the patient's autonomous mental processing. This conclusion is consistent with the finding that interpretations, which are potentially the most incisive interventions, had the strongest effects on alpha power. To our knowledge, this is the first attempt to examine if mutative psychoanalytic interventions, such as confrontations and especially interpretations, are associated with a modification of psychic processing in a patient as documented by the EEG.

Recent investigations with magnetic resonance imaging (MRI) and magnetoencephalogram (MEG) proved that the subcortical structures then interfere with cortical structures and show higher activity than during focused attention (Mazzetti et al., 2019). Neurophysiological studies showed that posterior neuronal oscillations in the alpha band (8–13 Hz) reflect the allocation of covert attention (Worden et al., 2000; Kelly et al., 2006; Thut et al., 2006; Jensen and Mazaheri, 2010) mediated by corticocortical interactions (Capotosto et al., 2012; Ptak, 2012; Vossel et al., 2014; Marshall et al., 2015a,b). Based on previous findings and theories that subcortical structures in the hypothalamus interfere with corticocortical neuronal structures during focused cognitive tasks and change the EEG rhythm (van Schouwenburg et al., 2010; Mazzetti et al., 2019), we may

hypothesize that mutative interventions in a patient with an impaired personality structure are accompanied by a change in psychic processing with a correlate in the EEG power spectrum.

The strength of our approach was to measure for the first time a patient's brain activity while being interviewed by an expert in the structural interview to identify expected differences in response to divergent psychoanalytic techniques. The main limitation of our study is that our findings are derived from one subject only. The results might be a very individual reaction of this one subject and dependent on age, gender, intellectual capacities of the subject, and other confounding variables like interaction with the interviewer and environmental factors. Moreover, we cannot completely rule out differences in motor behavior between the different conditions. It is possible that the patient was more agitated in one condition over the other and moved his leg(s) and/or hand(s) more relative to the control condition. This could possibly account for the reported differences observed over motor areas between the "interpretation" and "confrontation" and control conditions. Nevertheless, this single-case finding could serve as a first basis to repeat EEG recordings of a higher amount of psychoanalytic interviews with several subjects of different ages and gender. We may conclude that using EEG during therapeutic psychoanalytic techniques might be a helpful tool to evaluate differential responses to the process of psychodynamic psychotherapy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Review Board of the University of Innsbruck. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AB and KL conceptualized the study together with PB, BS-U, and OK. KL conducted the EEG assessment. KL and FB analyzed the EEG data and performed statistical analyses. OK administered the structural interview with the patient and evaluated the structural diagnoses of the patient. AB and PB rated and segmented the episodes (clarification, confrontation, and interpretation) of the structural interview and described the interview process of the patient. BS-U and OK described the clinical diagnosis of the patient. NN gave valuable expert support for interpreting the EEG results. CP-S gave valuable

support with respect to the EEG analyses. AB, KL, PB, and NN conceptualized the manuscript. AB and KL wrote the manuscript with contributions from all co-authors. All authors approved the final version of the manuscript and read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abbass, A., Nowoweiski, S., Bernier, D., Tarzwell, R., and Beutel, M. (2014). Review of psychodynamic psychotherapy: Neuroimaging studies. *Psychother. Psychosom.* 83, 142–147.
- Bateman, A., and Fonagy, P. (2004). Mentalization-based treatment of BPD. *J. Pers. Disord.* 18, 36–51.
- Bateman, A., and Fonagy, P. (2016). *Mentalization-based treatment for personality disorders – A practical guide*. Oxford: Oxford University Press.
- Beutel, M., Stark, R., Pan, H., Silbersweig, D., and Dietrich, S. (2010). Changes of brain activation pre-post short-term psychodynamic inpatient psychotherapy: An fMRI study of panic disorder patients. *Psychiatry Res. Neuroimaging* 184, 96–104. doi: 10.1016/j.pscychres.2010.06.005
- Beutel, M., Stark, R., Pan, H., Silbersweig, D., and Dietrich, S. (2012). Wie beeinflusst psychodynamische psychotherapie die hirnakktivierung von panikpatienten? Langzeitergebnisse einer fMRT studie. *Psychotherapeut* 57, 227–233.
- Blüml, V., and Doering, S. (2021). ICD-11 personality disorders: A psychodynamic perspective on personality functioning. *Front. Psychiatry* 12:654026. doi: 10.3389/fpsy.2021.654026
- Böker, H., Richter, A., Himmighoffen, H., Ernst, J., Bohleber, L., Hofmann, E., et al. (2013). Essentials of psychoanalytic process and change: How can we investigate the neural effects of psychodynamic psychotherapy in individualized neuroimaging? *Front. Hum. Neurosci.* 7:355. doi: 10.3389/fnhum.2013.00355
- Buchheim, A., Hörz-Sagstetter, S., Doering, S., Rentrop, M., Buchheim, P., Schuster, P., et al. (2017). Change of unresolved attachment in borderline personality disorder: RCT study of transference-focused psychotherapy. *Psychother. Psychosom.* 86, 314–316. doi: 10.1159/000460257
- Buchheim, A., Labek, K., Taubner, S., Kessler, H., Pokorny, D., and Kächele, H. (2018). Modulation of gamma band activity and late positive potential in patients with chronic depression after psychodynamic psychotherapy. *Psychother. Psychosom.* 87, 252–254. doi: 10.1159/000488090
- Buchheim, A., Viviani, R., Kessler, H., Kächele, H., Cierpka, M., Roth, G., et al. (2012). Changes in prefrontal-limbic function in major depression after 15 months of long-term psychotherapy. *PLoS One* 7:e33745. doi: 10.1371/journal.pone.0033745
- Buchheim, P., Cierpka, M., Kächele, H., and Jimenez, J. (1987). Das »strukturelle interview« ein beitrag zur integration von psychopathologie und psychodynamik im psychiatrischen erstgespräch. *Fundam. Psychiatr.* 1, 154–161.
- Caligor, E., Kernberg, O., Clarkin, J., and Yeomans, F. (2018). *Psychodynamic therapy for personality pathology: Treating self and interpersonal functioning*. Washington, DC: American Psychiatric Press.
- Capotosto, P., Corbetta, M., Romani, G., and Babiloni, C. (2012). Electrophysiological correlates of stimulus-driven reorienting deficits after interference with right parietal cortex during a spatial attention task: A TMS-EEG study. *J. Cogn. Neurosci.* 4, 2363–2371. doi: 10.1162/jocn_a_00287
- Carhart-Harris, R., and Friston, K. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
- Clarkin, J., Levy, K., Lenzenweger, M., and Kernberg, O. (2007). Evaluating three treatments for borderline personality disorder: A multiwave study. *Am. J. Psychiatry* 164, 922–928. doi: 10.1176/ajp.2007.164.6.922
- Clarkin, J., Yeomans, F., and Kernberg, O. (1999). *Psychotherapy for borderline personality. Focusing on object relations*, 2 Edn. Arlington, VA: American Psychiatric Publishing.
- Cohen, M. (2017). Where does EEG come from and what does it mean? *Trends Neurosci.* 40, 208–218. doi: 10.1016/j.tins.2017.02.004
- de Greck, M., Bölter, A. F., Lehmann, L., Ulrich, C., Stockum, E., Enzi, B., et al. (2013). Changes in brain activity of somatoform disorder patients during emotional empathy after multimodal psychodynamic psychotherapy. *Front. Hum. Neurosci.* 7:410. doi: 10.3389/fnhum.2013.00410
- de Greck, M., Scheidt, L., Bölter, A. F., Frommer, J., Ulrich, C., Stockum, E., et al. (2011). Multimodal psychodynamic psychotherapy induces normalization of reward related activity in somatoform disorder. *World J. Biol. Psychiatry* 12, 296–308.
- Doering, S., Hörz, S., Rentrop, M., Fischer-Kern, M., Schuster, P., Benecke, C., et al. (2010). Transference-focused psychotherapy v. treatment by community psychotherapists for borderline disorder, randomized controlled trial. *Br. J. Psychiatry* 196, 389–395. doi: 10.1192/bjp.bp.109.070177
- Etchegoyen, R. (1999). *The fundamentals of psychoanalytic technique*. London: Routledge.
- Fischer-Kern, M., Doering, S., Taubner, S., Hörz, S., Zimmermann, J., Rentrop, M., et al. (2015). Transference-focused psychotherapy for borderline personality disorder: Change in reflective function. *Br. J. Psychiatry* 207, 173–174.
- Gelman, A., and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Hager, B., Yang, A., and Gutsell, J. (2018). Measuring brain complexity during neural motor resonance. *Front. Neurosci.* 12:758. doi: 10.3389/fnins.2018.00758
- Høglend, P., Bøgwald, K., Amlø, S., Marble, A., Ulberg, R., and Sjaasrød, M. (2008). Transference interpretations in dynamic psychotherapy: Do they really yield sustained effects? *Am. J. Psychiatry* 165, 763–771. doi: 10.1176/appi.ajp.2008.07061028
- Høglend, P., Dahl, H., Hersoug, A., Lorentzen, S., and Perry, J. (2011). Long-term effects of transference interpretation in dynamic psychotherapy of personality disorders. *Eur. Psychiatry* 26, 419–424. eurpsy.2010.05.006 doi: 10.1016/j
- Hörz-Sagstetter, S., Caligor, E., Preti, E., Stern, B., De Panfilis, C., and Clarkin, J. (2018). Clinician-guided assessment of personality using the structural interview and the structured interview of personality organization (STIPO). *J. Pers. Assess.* 100, 30–42. doi: 10.1080/00223891.2017.1298115
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Kandel, E. (1998). A new intellectual framework for psychiatry. *Am. J. Psychiatry* 155, 457–469.

- Kandel, E. (2013). The new science of mind and the future of knowledge. *Neuron* 80, 546–560. doi: 10.1016/j.neuron.2013.10.039
- Kehyayan, A., Best, K., Schmeing, J., Axmacher, N., and Kessler, H. (2013). Neural activity during free association to conflict-related sentences. *Front. Hum. Neurosci.* 7:705. 2013.00705 doi: 10.3389/fnhum
- Kelly, S., Lalor, E., Reilly, R., and Foxe, J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *J. Neurophysiol.* 95, 3844–3851. doi: 10.1152/jn.01234.2005
- Kernberg, O. (1981). Structural interviewing. *Psychiatry Clin. North Am.* 4, 169–195.
- Kernberg, O. (1984). *Severe personality disorders. Psychotherapeutic strategies.* New Haven: Yale Univ Press.
- Kernberg, O. (1999). Psychoanalysis, psychoanalytic psychotherapy and supportive psychotherapy: Contemporary controversies. *Int. J. Psychoanal.* 80, 1075–1091. doi: 10.1516/0020757991599205
- Kernberg, O. (2016). What is personality? *J. Pers. Disord.* 30, 145–156. doi: 10.1521/pedi.2106.30.2.145
- Kernberg, O., Selzer, M., Koenigsberg, H., Carr, A., and Appelbaum, A. (1989). *Psychodynamic psychotherapy of borderline patients.* New York, NY: Basic Books.
- Klimesch, W., Doppelmayr, M., Schimke, H., and Pachinger, T. (1996). Alpha frequency, reaction time, and the speed of processing information. *J. Clin. Neurophysiol.* 13, 511–518.
- Levy, K., Keefe, J., and Ehrenthal, J. (2019). “Research support for psychodynamic constructs,” in *Contemporary psychodynamic psychotherapy: Evolving clinical practice*, eds D. Kealy and J. S. Ogrodniczuk (Cambridge, MA: Elsevier Academic Press), 89–106. doi: 10.1016/B978-0-12-813373-6.00006-4
- Levy, K., Meehan, K., Kelly, K., Reynoso, J., Weber, M., Clarkin, J., et al. (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *J. Consult. Clin. Psychol.* 74, 1027–1240. doi: 10.1037/0022-006X.74.6.1027
- Marshall, T., O’Shea, J., Jensen, O., and Bergmann, T. (2015a). Frontal eye fields control attentional modulation of alpha and gamma oscillations in contralateral occipitoparietal cortex. *J. Neurosci.* 35, 1638–1647. doi: 10.1523/JNEUROSCI.3116-14.2015
- Marshall, T., Bergmann, T., and Jensen, O. (2015b). Frontoparietal structural connectivity mediates the top-down control of neuronal synchronization associated with selective attention. *PLoS Biol.* 13:e1002272. doi: 10.1371/journal.pbio.1002272
- Mazzetti, C., Staudigl, T., Marshall, T., Zumer, J., Fallon, S., and Jensen, O. (2019). Hemispheric asymmetry of globus pallidus relates to alpha modulation in reward-related attentional tasks. *J. Neurosci.* 39, 9221–9236. doi: 10.1523/JNEUROSCI.0610-19.2019
- Messina, I., Sambin, M., Beschner, P., and Viviani, R. (2016). Changing views of emotion regulation and neurobiological models of the mechanism of action of psychotherapy. *Cogn. Affect. Behav. Neurosci.* 16, 571–587. doi: 10.3758/s13415-016-0440-5
- Milrod, B., Busch, F., Cooper, A., and Shapiro, T. (1997). *Manual of panic-focused psychodynamic psychotherapy.* Washington, DC: American Psychiatric Press.
- Panksepp, J., and Solms, M. (2012). What is neuropsychanalysis? Clinically relevant studies of the minded brain. *Trends Cogn. Sci.* 16, 6–8. doi: 10.1016/j.tics.2011.11.005
- Perez, D., Vago, D., Pan, H., Root, J., Tiescher, O., Fuchs, B., et al. (2016). Frontolimbic neural circuit changes in emotional processing and inhibitory control associated with clinical improvement following transference-focused psychotherapy in borderline personality disorder. *Psychiatry Clin. Neurosci.* 70, 51–61. doi: 10.1111/pcn.12357
- Pfurtscheller, G., and Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857. doi: 10.1016/s1388-2457(99)00141-8
- Pincus, A., Cain, N., and Halberstadt, A. (2020). Importance of self and other in defining personality pathology. *Psychopathology* 53, 133–140. doi: 10.1159/000506313
- Pinheiro, J., and Bates, D. (2000). *Mixed-effects models in S and S-PLUS.* New York, NY: Springer. doi: 10.1007/b98882
- Prosser, A., Friston, K., Bakker, N., and Parr, T. (2018). A bayesian account of psychopathy: A model of lacks remorse and self-aggrandizing. *Comput. Psychiatry* 92:140. doi: 10.1162/cpsy_a_00016
- Ptak, R. (2012). The frontoparietal attention network of the human brain: Action, saliency, and a priority map of the environment. *Neuroscientist* 18, 502–515. doi: 10.1177/1073858411409051
- Schmeing, J., Kehyayan, A., Kessler, H., Do Lam, A., Fell, J., Schmidt, A., et al. (2013). Can the neural basis of repression be studied in the MRI scanner? New insights from two free association paradigms. *PLoS One* 8:e62358. doi: 10.1371/journal.pone.0062358
- Schmidt, R., Ruiz, M., Kilavik, B., Lundqvist, M., Starr, P., and Aron, A. (2019). Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function. *J. Neurosci.* 39, 8231–8238. doi: 10.1523/JNEUROSCI.1163-19.2019
- Shevrin, H., Snodgrass, M., Brakel, L., Kushwaha, R., Kalaida, N., and Bazan, A. (2013). Subliminal unconscious conflict alpha power inhibits supraliminal conscious symptom experience. *Front. Hum. Neurosci.* 7:544. doi: 10.3389/fnhum.2013.00544
- Solms, M. (2011). Neurobiology and the neurological basis of dreaming. *Handb. Clin. Neurol.* 98, 519–544. doi: 10.1016/B978-0-444-52006-7.00034-4
- Solms, M. (2021a). Revision of drive theory. *J. Am. Psychoanal. Assoc.* 69, 1033–1091. doi: 10.1177/00030651211057041
- Solms, M. (2021b). *The hidden spring: A journey to the source of consciousness.* London: Giuardian Books.
- Thomä, H., and Kächele, H. (1985). *Lehrbuch der psychoanalytischen therapie. Bd.1: Grundlagen.* New York, NY: Springer.
- Thut, G., Nietzel, A., Brandt, S., and Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26, 9494–9502. doi: 10.1523/JNEUROSCI.0875-06.2006
- Twisk, J. W. (2006). *Applied multilevel analysis: A practical guide for medical researchers.* Cambridge: Cambridge university press.
- Unterrainer, H., Chen, M., and Gruzelier, J. (2014). EEG-neurofeedback and psychodynamic psychotherapy in a case of adolescent anhedonia with substance misuse: Mood/theta relations. *Int. J. Psychophysiol.* 93, 84–95. doi: 10.1016/j.ijpsycho.2013.03.011
- Unterrainer, H., Lewis, A., and Gruzelier, J. (2013). EEG-Neurofeedback in psychodynamic treatment of substance dependence. *Front. Psychol.* 2013:692. doi: 10.3389/fpsyg.2013.00692
- van Schouwenburg, M., Aarts, E., and Cools, R. (2010). Dopaminergic modulation of cognitive control: Distinct roles for the prefrontal cortex and the basal ganglia. *Curr. Pharm. Des.* 16, 2026–2032. doi: 10.2174/138161210791293097
- Vossel, S., Geng, J., and Fink, G. (2014). Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist* 20, 150–159. doi: 10.1177/1073858413494269
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer.
- Williamson, S., Kaufman, L., Lu, Z., Wang, J., and Karron, D. (1997). Study of human occipital alpha rhythm: The alphon hypothesis and alpha suppression. *Int. J. Psychophysiol.* 26, 63–76. doi: 10.1016/s0167-8760(97)00756-3
- Worden, M., Foxe, J., Wang, N., and Simpson, G. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J. Neurosci.* 20:RC63. doi: 10.1523/JNEUROSCI.20-06-j0002.2000
- Yakeley, J. (2018). Psychoanalysis in modern mental health practice. *Lancet Psychiatry* 5, 443–450. doi: 10.1016/S2215-0366(18)30052-X



OPEN ACCESS

EDITED BY

Mark Leonard Solms,
University of Cape Town, South Africa

REVIEWED BY

Pierre-Henri Castel,
CNRS UMR 8065 Lier-FYT, France
Yasutaka Kubota,
Shiga University, Japan

*CORRESPONDENCE

Giulia Olyff
✉ giulia.olyff@ulb.be

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 09 June 2022

ACCEPTED 20 December 2022

PUBLISHED 10 February 2023

CITATION

Olyff G and Bazan A (2023) People
solve rebuses unwittingly—Both
forward and backward: Empirical
evidence for the mental effectiveness
of the signifier.
Front. Hum. Neurosci. 16:965183.
doi: 10.3389/fnhum.2022.965183

COPYRIGHT

© 2023 Olyff and Bazan. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

People solve rebuses unwittingly—Both forward and backward: Empirical evidence for the mental effectiveness of the signifier

Giulia Olyff^{1,2*} and Ariane Bazan^{1,2,3}

¹Centre de Recherche en Psychologie Clinique, Psychopathologie et Psychosomatique, Université libre de Bruxelles (ULB), Bruxelles, Belgium, ²Observatoire du SIDA et des Sexualités, Université libre de Bruxelles (ULB), Bruxelles, Belgium, ³Interpsy (UR 4432), Université de Lorraine, Nancy, France

Introduction: Freud proposed that names of clinically salient objects or situations, such as for example a beetle (*Käfer*) in Mr. E's panic attack, refer through their phonological word form, and not through their meaning, to etiologically important events—here, “*Que faire?*” which summarizes the indecisiveness of Mr. E's mother concerning her marriage with Mr. E's father. Lacan formalized these ideas, attributing full-fledged mental effectiveness to the signifier, and summarized this as “the unconscious structured as a language”. We tested one aspect of this theory, namely that there is an influence of the ambiguous phonological translation of the world upon our mental processing without us being aware of this influence.

Methods: For this, we used a rebus priming paradigm, including 14 French rebuses, composed of two images depicting common objects, such as *paon* /pā/ “peacock” and *terre* /tɛr/ “earth,” together forming the rebus *panthère* /pātɛr/ “panther.” These images were followed by a target word semantically related to the rebus resolution, e.g., *félin* “feline,” upon which the participants, unaware of the rebus principle, produced 6 written associations. A total of 1,458 participants were randomly assigned either to Experiment 1 in which they were shown the rebus images in either forward or in reverse order or to Experiment 2, in which they were shown only one of both rebus images, either the first or the last.

Results and discussion: The results show that the images induced inadvertent rebus priming in naïve participants. In other words, our results show that people solve rebuses unwittingly independent of stimulus order, thereby constituting empirical evidence for the mental effectiveness of the signifier.

KEYWORDS

rebus, signifier, Freud, Lacan, phonological priming, Shevrin

*In fond memory of Howard Shevrin (1926-2018)
careful reader of Freud and revolutionary scientist of the human mind.*

1. Introduction

Lacan (1966, p. 868) proposed that “the unconscious is structured as a language” and in previous theoretical work, elaborated in *“Phantoms in the voice. A neuropsychanalytic hypothesis on the structure of the unconscious”* (Bazan, 2007), the second author spelled out an interdisciplinary neuropsychanalytic framework for this claim. This research now proposes an experimental testing for one aspect of Lacan’s proposal, namely that the world is also apprehended as a phonological trace, independently of its semantic meaning. To show this, we have devised a supraliminal rebus priming paradigm in which two images were presented side by side (e.g., an image of *paon* /pā/ “peacock” and of the *terre* /tēr/ “earth”), which together form a new French word or rebus (namely *panthère* /pāter/ “panther”). After looking to the images for 4s, the participants were asked to write their first 6 associations to a target word, semantically related to the rebus resolution (e.g., *félin* “feline”). If there were more rebus resolution words (in the given example, *panthère*) in these associations than there were in a control priming, in participants who stayed naive to the protocol, then it was considered that people can solve rebuses unwittingly, thereby constituting evidence going into the direction of Lacan’s proposition.

In the fifties of the previous century, Lacan (1955) operates what is called a “return to Freud”. One of his major concerns in doing so, is that scholars after Freud read his texts on the semantic interpretational level. Indeed, Freud’s major interpretation myths, such as the Oedipus and the Narcissus myth, the Myth of Primitive Hord, and the castration anxiety, are in a way so mind-blowing that they easily take up the whole mental space. We are left blinded by these storylines with echoes all through mythology, literature, and the arts, and thereby, we tend to overread or neglect another major discovery. Freud indeed reveals another shocking new take on how humans grasp the world. First, we must acknowledge that the reigning cognitive paradigm for the way in which humans grasp the world is overwhelmingly visual-semantic (see e.g., Kosslyn, 1994): this is, humans tend to understand the world—the external as well as the internal imagery world—on the semantic interpretation of its imagery appearance. Freud, however, understands that another logic, a phonological one, also underlies our mental grasping of things. In a letter to his friend Wilhelm Fliess, on 29 December 29 1897 (Freud, 1897), he describes the case of Mr. E. This patient evokes his panic attacks as a ten-year-old when trying to catch a black beetle, a *Käfer* in German. It is Mr. E himself who in the session reveals the mental meaning of this black beetle, by shifting the phonological reading from *Käfer* to the French *Que faire?* which, when pronounced with a German accent, sounds about the same way (Mr. E in fact learned French before learning German with his French nanny). However, *Que faire?* means “What to do now?”, a key phrase reflecting both Mr. E’s

central symptom—his indecisiveness—and one of the probable etiological origins of his distress, namely the inability of his mother to make up her mind concerning her marriage—which concerns, of course, the choice for Mr. E’s father. At the end of this letter, Freud adds the Yiddish expression *Meschugge!* approximately meaning “isn’t this completely crazy?”. But as a true scientist, Freud does not discard the parts of reality which do not please him or which do not work with his (rational) vision of the human condition. Instead, he starts to gather more materials of the phonological logic underlying mental symptoms, such as everyday life accidents (the forgetting of words, e.g., *Signorelli* in *Psychopathology of Everyday life*; Freud, 1901) and obsessions (e.g., the rat-obsession in *The Ratman*; Freud, 1909). He even notices it in non-pathological preferences: “a young man had exalted a certain sort of ‘shine on the nose’ into a fetishistic precondition. The surprising explanation of this was that the patient had been brought up in an English nursery but had later come to Germany, where he forgot his mother-tongue almost completely. The fetish, which originated from his earliest childhood, had to be understood in English, not German. The ‘shine on the nose’ [in German ‘*Glanz auf der Nase*’] was in reality a ‘glance at the nose.’ The nose was thus the fetish, which, incidentally, he endowed at will with the luminous shine which was not perceptible to others.” (Freud, 1927/1931; p. 152) and in disgusts: “A young lady can’t retrieve the name of the Lewis Wallace novel ‘Ben Hur’. When she analyses this forgetting, she realizes in the aftermath that it is because of its phonological closeness to *bin Hure* ‘I am a whore’ which, as she says: « contains an expression that I do not care to use, especially in the company of young man” (Freud, 1901, p. 41). Mostly such a rebus structure of our ‘symptoms’ remains obscure, even to ourselves, but in some personality structures the unconscious is at the surface and rebus reading is mobilized consciously. What is common to all these symptomatic expressions is that phonological ambiguity is mobilized (unconsciously or consciously) to operate drastic shifts in semantic realms, sometimes directly expressed in behavior (e.g., the preference for a shine on the nose), obscuring the mediating role of language. There are many other examples throughout Freud’s whole *oeuvre*, in the first place in *The Interpretation of Dreams* (Freud, 1900).

When it comes to dreams, Freud (1900, p. 242), is indeed elated when he says: “... a dream is [such] a picture-puzzle,¹ and our predecessors in the art of dream-interpretation have made the mistake of judging the rebus as an artistic composition. As such, of course, it appears nonsensical and worthless.” Several examples of such rebus readings in dreams were given before (Steinig et al., 2017, pp. 2–3). To repeat a straightforward one, in an example of the second author’s practice, a woman, upon coming back from South Africa, dreamt she was paying in a

1 or *rebus* and translated as “rebus” in the French translation. In the original German version: “Ein solches Bilderrätsel ist nun der Traum”.

bar with white pieces of paper upon which a big square was drawn almost filling the square. The figurative value of this dream leaves us clueless. But, when she recounts the dream, she indicates she was paying with pieces of paper on which “rand”-s were drawn, *rand* being both the Dutch word for edge and for the South African currency: the rebus reading reveals the meaning. In a new example, a patient dreamt that people were running 78 times back-and forwards on a sand beach, and this dream made sense with the simple intervention “78 rotations per minutes,” or 78 RPM—the dreamer thereupon remembering an emotional childhood scene with a DJ playing records on a vacation beach. Judging the dream “as an artistic composition,” i.e., on the face validity of the image, would have left us stuck on the back-and forth running with a mysterious 78—possibly leading to mystical interpretations of the “magical 7” followed by the “eternal 8.” The rebus interpretation, which arises from the way in which the analysand addresses the dream, on the other hand, gives access to one precise memory. In other words, it is when we read the dream as a picture-puzzle or a rebus, that it delivers (some of) its underlying secrets.

Common to these diverse examples is thus the observation that it is through the phonological structure of the involved name—and not primarily through the semantic signification of the involved “thing”—that key element in symptoms refer to the subject’s history and suggest a plausible rationale for the symptom or the inclination of the subject. Freud even explains how such a “background” network of singular intensities around phoneme groups influences our word choices unconsciously when we speak: “We think we are generally free to choose words and images to express our ideas. But a closer observation shows that it is often considerations extraneous to the ideas that decide this choice and that the form in which we mold our ideas often reveals a deeper meaning, which we do not realize ourselves. (...) some of these images and ways of speaking are often allusions to subjects which, while remaining *in the background*, exert a powerful influence on the speaker. I know someone who, at one time, continuously used, (...) the following expression: ‘When something suddenly crosses the head of someone.’ Now I knew that the man who spoke in this way had recently received the news that a Russian projectile had passed through the field-cap which his son, a fighting soldier, had on his head.” (Freud, 1901, p. 239; our translation and our Italics) and we have proposed a neuroscientific explanation for these observations (Bazan, 2007, 2011; Bazan et al., 2019).

It might seem quite a challenge to find a way to put this proposition to the test and we do not pretend we propose here its full operational equivalent. However, the phonological symptom logic, shown above, supposes that the surrounding world, even when we do not specifically convene it linguistically, has nevertheless a mental influence upon us through its specifically linguistic structure. Importantly, the specifically linguistic influence can only be revealed as a ‘phonological’ influence: indeed, if an exclusively semantic influence was

shown, it might be disputed that it should obligatorily be a linguistic one, as it might have been understood as an experiential or imagery influence, translated secondarily in words. But for example, the black beetle is supposed to induce anxiety upon Mr. E. not solely through the black, erratic appearance of its “thing,” the animal beetle, but (importantly) through the linguistic structure of its name, *Käfer*, referring to the anxiogenic ambivalent choice of Mr. E’s mother for Mr. E’s father: *Que faire?* (“What to do now?”). But also, the shiny noses of women are supposed to induce excitement upon the young English-German patient through the phonological equivalence between *Glanz* and glance and the Ben Hur book title is thought to induce aversion through the phonological closeness with *bin Hure*. Here is another example entrusted to the second author: “*Amici* is the name of a shop I once went to with my parents when I was thirteen. I never forgot the name and I also really wanted a coat from that store. Once in India it suddenly turned out that our old caretaker of the orphanage was called ‘*Amaci*.’ That was her nickname and means ‘mother.’ We were adopted at 11 months”. Here, the shop, in a rebus manner, translates to the mothering caretaker, affecting the mental realm of the teller. In other words, even when we do not consciously speak the word, an unconscious influence of its phonological translation upon our mental processing is supposed.

It is therefore disconcerting to us, as it was to Lacan (1955), to observe that the structural importance of language in organizing unconscious mental life, is largely dismissed in modern psychoanalysis. Though in our reading, Freud does his utmost best, on almost every page of the *Interpretation of dreams* (1901) for example, to underscore the formally linguistic weave of dreams, it comes to us that modern psychoanalysis and neuropsychanalysis, classify (away) linguistic phenomena as conscious, secondary, even cognitive, processes and do not consider linguistic structure as constitutive of unconscious mental life, despite the many examples all through Freud’s writings. For example, we deplore the way “signifiers” or “word presentations” are sometimes classified among other non-linguistic “signifiers” in modern psychoanalysis. This, to us, is a more subtle way of doing away with the radical consequences of the constitutive role of signifiers, namely that our lives and “personalities” are also influenced by coincidental forms which do not follow a semantic, rational logic but operate upon all of us in an illogical, crazy—*Meschugge!*—way. Whatever other motor species one would wish to classify under “signifiers,” there is no motor sequence of any other kind which can switch radically from semantic realm while remaining mechanically completely identical, i.e., in the precise physiology and biomechanics of the involved muscles and joints: this is the unique characteristic of language and this has consequences for our mental constitution, as shown in the clinical examples.

It is also this aspect—the influence of the phonological translation of the world upon our mental processing—that we are putting to the test in this research. For doing this, we are

using a rebus paradigm. Indeed, in a rebus the image does not replace a word but the sound—or even better the *phonology* or articulation (Bazan, 2007)—of that word (Préaud, 2004). A rebus here is constituted of two images denoting things, which, translated into their names and put side by side, form a new word, with a radically different meaning. For example, Shevrin and Luborsky (1961)'s most well-known rebus in their early dream-rebus studies is composed of the image of a pen, followed by the image of knee (see also Figure 1), resulting in the rebus 'penny'. When these images were flashed for 6 ms before sleeping, this led to more 'penny' associates (e.g., coin, money) than the control rebus when people were awakened in their REM phase of dreaming. This research was recently replicated by the German researcher Steinig et al. (2017) with the *kampflos* rebus—meaning “without a fight” in German—formed by the image of a comb (*Kamm*)— and of a raft (*Floß*; see Figure 1), leading to essentially the same results.

What this research shows and what ours propose to test is the general “rebus-functioning” mode of the human mental assessment of the world, as illustrated by the clinical examples, more than the empirical confirmation of a clinically observed rebus-resolution instance. Indeed, it is noteworthy that even if “the dream is a rebus” is a key phrase in Freud's oeuvre, we have found no example of a real rebus or word-puzzle, involving the combination of the names of two or more pictorially represented objects, in the whole of Freud's writings. As shown, there are many examples involving the necessity to translate an object, a disposition, or an action into language, to obtain the key to a dream or a symptom, but none of them involving a combination of several juxtaposed images—except in Freud's (1900, p. 242) theoretical example: “a house with a boat on its roof, a single letter of the alphabet, the figure of a running man whose head has been conjured away,” even if, frustratingly, he does not give the resolution of this rebus. Of course, it is fair to suppose that rebus resolution as the phonological translation of two or more juxtaposed images indeed as such also happens in dreams and symptomatic formations, but is less easy to detect, to start by the subject him or herself. That this supposition is not out of hand is confirmed by the fact that others, as Shevrin and Luborsky (1961), and Steinig et al. (2017), have had the same bet. More fundamentally, even in the case that this type of rebus resolution would *not* happen, if such clinical observations as listed here are true, then the unconscious transformation of images to phonological traces should be possible (see also Shevrin and Fritzler, 1968). When two images are given side by side, if they are effectively phonologically translated and this phonological read-out has a form of independence to switch to other semantic realms, then this should be detectable in the slightly higher probability of rebus resolution. For this reason, the rebus paradigm, if it works, does suppose the phenomena which we want to show and contributes to the idea of a generally human rebus-mode of grasping the world.

There is, however, a big difference between the former dream-rebus research and the present one: we have not put ourselves in circumstances in which an increased unconscious functioning is supposed nor induced—we have not worked with dreams, nor with subliminal presentations. Indeed, we presumed that rebus resolution should not only be measurable in subliminal paradigms or when awakened during dreaming but also—though probably only very slightly (see further)—in supraliminal conditions in fully awake nonclinical participants. Therefore, we designed a rebus paradigm consisting of 14 French rebuses, composed of two images depicting monosyllabic words of easily identifiable common objects, such as e.g., *panthère* /päter/ “panther,” composed by the images of *paon* /pā/ “peacock” and of *terre* /tēr/ “earth.” We presented these rebus images in a priming paradigm, i.e., the images were followed by a word related to the rebus resolution, e.g., *félin* “feline”; this is the rebus resolution, and the target words are semantic neighbors (such as, in this case, “panther” and “feline”). The participants were asked to look at the images for four seconds and then to write down the six first associations to the prime word. However, they were not specifically told to do something with the images, let alone that the images were forming a wordplay. The same prime words were also preceded by an unrelated rebus formed by two images, with the same instruction for six associations, constituting the inter-subject control situation. If more rebus-resolutions were counted among the associations upon the prime word, when this was preceded by its rebus than when it was preceded by an unrelated rebus, an influence of the rebus upon the associations must be supposed.

This, then, is our hypothesis, namely that *a pictorial environment can prime its reading on a phonological level—and thus, specifically, that two depicted images can prime their rebus reading—in a situation where the images were not convened linguistically*. This kind of passive phonological priming by pictorial materials was studied before, although in substantially different paradigms. For example, Humphreys et al. (2010) asked participants to associate upon a target word (e.g., cobweb) that was presented together with an image depicting an object with common phonemes in its name (here: the image of a spoon for the resolution word “spider”); participants were instructed to ignore the picture. They found more and faster spider-associations when the image of a spoon was presented suggesting that the common phoneme in “spoon” and “spider” had a facilitation effect in the association upon the target word “cobweb.” Studying cognitive and perceptual processes, Chabal and Marian (2015) test how overlapping picture names can influence visual-search performance: concretely, these participants are preferentially drawn to look at an image of a *cloud* in a picture (among other competing images) when this was preceded by the passive viewing of the picture of a *clock*, the effect being mediated by the common phonemes in the names of both items, even if no language was convened. Their conclusion



supports the idea of an “automatic language activation during visual processing” (p. 548) even when language is not introduced in the task. These results are in line with other studies investigating how phonologically related pictures can facilitate the spoken-word production in picture naming tasks. For example, participants were asked to name the line-drawing of a bed (in green) superposed with one of a bell (in red) in the phonologically related, or with one of a hat (in red) in the unrelated distractor condition; participants were faster to name the bed when superposed with the image of a bell (Morsella and Miozzo, 2002). Others (e.g., Navarrete and Costa, 2005; Huettig and McQueen, 2007; Roelofs, 2008) have replicated this paradigm, including with briefly flashed action scenes (e.g., of diving or *tauchen*) priming for the naming of action scenes with rhyming names (e.g., *rauchen*, to smoke; Zwitserlood et al., 2018). These studies show the “automatic” effect of the visual environment upon the mental activity of participants subjected to this environment. Nevertheless, only one of these studies involved the reemergence of the complete name of the depicted object in the subject’s speech with a radically different meaning: Meyer and Damian (2007) show that people are faster to name the image of e.g., a bat (baseball) when superposed by a drawing of a bat (animal). This study is the closest to the clinical situation, where similar meaning reversals are observed (such as the squared *rand*-drawing stood for the *rand*-South African currency in the dream excerpt cited earlier). However, clinically, often the shifts in meaning are more spectacular with a tipping point similar in phonology even if quite dissimilar in orthography (such as *Käfer/Que faire?* and *comme il serre/commissaire*—see further) and probably therefore often unnoticed. Our research is especially interested in these kinds of shifts.

We have avoided the words “passive” or “automatic” to indicate the absence of the need to convene the images linguistically for there to be an effect. In a Freudian perspective, mental processing may go unnoticed by the subject himself without excluding subject specific influences; these processes are called “primary processes”. Indeed, the so-called primary

process mental dynamic (Freud, 1900; see also Shevrin and Luborsky, 1961, p. 480) is an essentially associative mental dynamic, whereby connections are established based on so-called “superficial” (Freud, 1900, p. 597) or “non-essential” (Freud, 1905, p. 88; Holt, 1967, p. 354) characteristics, or even “attributes” (Rapaport, 1951, p. 708). Primary process mentation is thought to prevail in unconscious processing (Freud, 1915) but is also part of conscious mental dynamics. Primary processes function on a mirror-plane level with the stimulus in a direct action-reaction pattern, while for the secondary process to happen a third point, a perspective, is needed (Bazan, submitted). Secondary process mentation, to the contrary, keeps in mind the intentional structure of human reality and by doing so, can inhibit primary process connectivity (Freud, 1895, p. 334). For this reason, we will prefer the expressions “primary process” or “inadvertent” rebus priming, i.e., not achieved through deliberate planning. Moreover, Freud (1915, p. 186) said “By the process of displacement one idea may surrender to another its whole quota of cathexis; by the process of condensation, it may appropriate the whole cathexis of several other ideas. I have proposed to regard these two processes as distinguishing marks of the so-called primary psychical process”. As rebus resolution supposes the displacement of the cathexis to an attribute of the depicted object—its name—as well as its condensation with another name, rebus resolution without conscious injunction to resolve a rebus, is thought of as a primary process by excellence. Specifically, Shevrin et al. (1996, p. 107) claim that rebuses investigate “a formal aspect of dynamic unconscious thought organization marked primarily by superficial associations in the form of phonetic transitions and combinations,” which characterizes the primary process (Steinig et al., 2017). Importantly, primary process mentation is not concerned with spatial orientation: indeed, it is characterized by spatiotemporal confusion (Freud, 1900, 1936; Lacan, 1964, p. 40; Kapsambelis, 2005, p. 61) while spatiotemporal distinctions are, on the other hand, typical for secondary process mentation (Bazan, 2007). If rebus resolution is thought to happen along primary process logics, then people

should resolve them as easily when the images are given in a forward order (*paon* /pā/ “peacock” + *terre* /tɛr/ “earth”) and when they are given in the reverse order (*terre* /tɛr/ “earth” + *paon* /pā/ “peacock”). Therefore, we included both a forward and a reverse condition in an intersubject set-up in Experiment 1. We could not risk intrasubject testing, as people are more easily inclined to consciously understand the rebus principle in the forward than in the reverse condition, and therefore the forward condition would have “contaminated” the reverse condition.

We remained with one concern, even if not a prohibitive one. Indeed, if priming by one image forming the rebus was also resulting in significant rebus resolution, then it remains in this set-up impossible to hold that the rebus solutions were to be explained obligatorily by the simultaneous depicting of two images and the subsequent condensation of their two names. Still, in that case, positive rebus resolution results would have corroborated our first hypothesis of a mental influence of the linguistically translated environment. But only if we would be able to show significantly *more* rebus resolution when prime words were preceded by the two images forming the rebus, and significantly less (or else, *no* significant) rebus resolution when the prime word was preceded by only one image, could we suppose that there indeed was inadvertent *rebus* priming. In our theoretical psychoanalytical framework, which implies supposedly complex mental unconscious processing, it is more convincing of the unsuspected capacities of the unconscious to show rebus resolution than to show ‘simple’ phonological priming by pictorial material. For this reason, we also included a second experiment, Experiment 2, in our research, implying the priming by only one of the two images, either the first or the last, in our set-up in an intersubject-configuration—as, again, in an intrasubject configuration with both one and two-image presentations, the complete rebus trials would have contaminated the one-image trials by enhancing insight into the principles underlying the protocol. Only if there is less, or else no significant, rebus resolution in Experiment 2 (one image) as compared to Experiment 1 (two images) our hypothesis of unwitting *rebus* resolution can be confirmed.

However, showing one-image phonological influence would still confirm the possibility of phonological priming by pictorial stimulus material. Phonological priming by linguistic stimulus material is, of course, a well-known effect very largely used to study naming processes (Meyer et al., 1974; Emmorey, 2002). However, this is explicitly not the present paradigm, which works with pictorial priming material, interrogating its phonological, not semantic, influence. But, of course, a conscious effort to retrieve the names of the presented pictures would transform the experiment into a “classical” case of phonological priming. Therefore, to avoid indirect phonological priming by linguistic primes, or, in psychoanalytic terms, to remain in a primary process mental logic, it is important to exclude the rebus resolution in participants who understood the

rebus principle at some point during the research. Indeed, in that case, a conscious or attention-requiring effort to retrieve the names of the presented pictures may not be excluded. However, as will be shown, several people did indeed understand that the images were forming rebuses. For this reason, we had a thorough debriefing on people’s naivety at the end of the research and included only naive participants in all our analyses.

In sum, what we propose is that in any encounter of a subject with the visual world there will be a very marginal, though significant, rebus effect of this visual world upon the mental life of this subject. However, we only expect minimal rebus-effect as the mental influence of pictorial material is expected to be quite overwhelmingly visual (e.g., Kosslyn, 1994). Our proposition is that next to this well-known cognitive visual-semantic main effect, images can also prime for a different semantic realm than by the directly depicted one (the visuo-semantic translation) and this by way of their phonological translation. Nevertheless, if we want to make sense of the existing many clinical observations, this small effect should be significantly measurable even in these supraliminal conditions. Therefore, the present research thus proposes to verify if surrounding images, independently of their spatial configuration, have an inadvertent influence upon our mental productions through the phonology of their names and, moreover, that people produce rebus associations, independent of the order of image presentation, and significantly more so than by phonological association upon one image only. In other words, we propose to show that people solve rebuses unwittingly.

2. Experiment 1: Rebus priming

2.1. Materials and methods

2.1.1. Participants

Participants were recruited on social networks, and all were majors. A total of 906 participants completed the entire online survey hosted by the Limesurvey platform (Limesurvey, 2020). Of the 906 participants, 118 participants were eliminated. First, 43 answered “yes” or “don’t know/don’t remember” to the question: “Do you remember having formerly participated in a study by Giulia Olyff?” and were consequently discarded. This question was asked at the beginning and at the end we reiterated the question “Do you remember having participated in a similar study?”. There were 15 pilot studies before this one, and it was important to avoid participants who were already exposed to rebus stimuli and might therefore be less naive to the rebus principle. Second, 75 participants were eliminated because they were not native French speakers or because they had a self-assessment of French language proficiency ≤ 5 on a 7-points Likert scale. French language proficiency was tested by an objective measure on 12 points, the FR12, which is a quick French test, which we designed for this study

TABLE 1 Demographic data: Means (*M*) and standard deviations (*SD*) for *N* = 788 participants of Experiment 1.

	<i>M</i> (<i>SD</i>)
Age (years)	31.2(12.5)
Education (years)	15.5(1.8)
Self-assessment French proficiency (/7)	6.9(0.1)
French proficiency test (/12)	11.0(1.1)
Gender (% women)	76.6
Handedness (% right-handed)	84.5

(see **Supplementary material 1**). After informed consent, our participants were randomly assigned to one of the two variations of the protocol (see further section “2.1.3 Conditions, variations, and randomization”). All demographic data are available in **Table 1**.

2.1.2. Stimuli

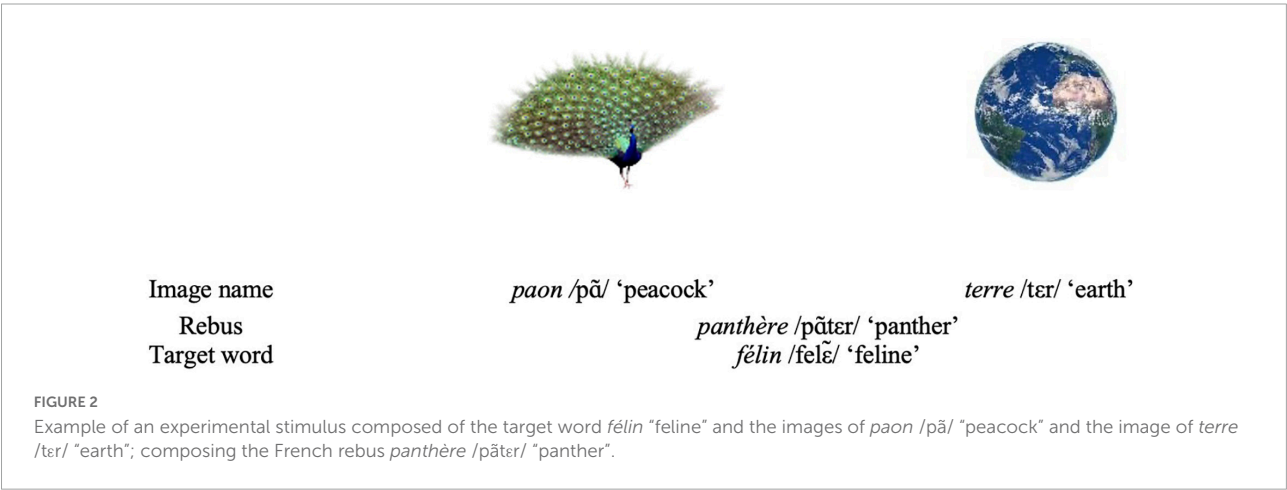
A stimulus is composed of a target word preceded by two images, forming a rebus. We designed 21 French rebuses composed by two images side by side. Each image represents a French monosyllabic word. Fourteen rebuses have an associated target word and together they compose our 14 stimuli. The remaining 7 rebuses that have no target word associated are our “supplementary rebuses” and are used for the CR condition (see further). For an example, see **Figure 2**, for the full list of stimuli see **Supplementary material 2**.

A total of 15 pilot studies, involving 3,327 participants, were conducted to develop our rebus priming method, including the selection of the rebuses, the rebus images, and the target words. To design our rebuses, we choose disyllabic French words of which the phonological translation can be divided into two monosyllabic French words. As a first requirement, the phonology of these words (forming the rebus) and the rebus in and by itself had to be perfectly identical. For example, *phobie* /fobi/ “phobia,” can be a rebus composed by the images of a

faux /fo/ “scythe” and of a *bille* /bij/ “ball” but the match of the phonological transcription of the two images and the rebus resolution word was not perfect and the rebus was discarded (/fobi/ and /fobij/). Second, the imaged words and their rebuses must have (radically) different etymologies. For example, the rebus word *muraille* /myraj/ “wall” may not be formed by images of *mur* /myr/ “wall” and *ail* /aj/ “garlic,” as the etymology of *muraille* refers to *mur*. Therefore, while being homophonic, all the imaged words and the syllables of their rebuses were nevertheless orthographically dissimilar. In the cited example, we used the images of *mûre* /myr/ “blackberry” and *ail* /aj/ “garlic.” As a third requirement for the rebuses, they had to be composed by imageable, easy-to-name rebus components. For example, *baleine* /balèn/ “whale,” is a great rebus, formed by *bas* /ba/ “stocking,” and *laine* /lèn/ “wool,” but both images are hard to picture in a recognizable way (stockings are difficult to picture in an unambiguous way, it turned out we had a lot of confusion between stockings, pantyhose, and socks).

For the images, we kept colored pictures with no background chosen among copyright-free images on the internet with the best naming percentages (correct naming: *M* = 85.2%, *SD* = 10.56; for the list of images see **Supplementary material 2**).

For the target words, we use semantic neighbors. We selected words that were chosen to be at maximum slightly associated to the rebus. Indeed, highly associated target words would give high “rebus resolution” even in the control condition. For an objective measure of this association, we produced associative French norms to our target words (Olyff et al., submitted). The associative strength range between the target word and the rebus resolution was between 0 and 8.7%. These percentages indicate that, for example, when we presented the word *fromage* “cheese,” 4% of participants give as *first* association the word *dessert* in no-priming set-ups. The associative strength percentages predominantly serve to avoid highly associated target words. Even though the associative strength may be 0% for the first association, words



are nevertheless chosen in association to the target in further associations. For our 14 stimuli the percentage associative strength between the target word and the rebus resolution word had a $M = 1.8$, $SD = 2.73$; for the list of associative strength between rebus resolution word and target word see [Supplementary material 2](#).

2.1.3. Conditions, variations, and randomization

Each target word could be presented in two conditions: the experimental condition (EX) or the control condition (CR). The target words in the CR condition must control for the spontaneous priming of the rebus resolutions, even without the related rebus images and in this case, with unrelated control rebus images. For the control condition we used unrelated rebuses because we cannot use rebus images of the experimental condition. Indeed, preliminary studies have shown that pseudo-randomizing target words with rebuses from the pool of experimental rebus image couples to obtain the control condition, gives rise to unwanted late priming effects. Consider e.g., the following situation: the images *rape* /rap/ “grater” and *as* /as/ “ace” used as control rebus images are followed by an unrelated target word; somewhat later another control stimulus is followed by an unrelated target word *aigle* “eagle”; though the images and the target words are mismatched, the coincidental order might still enhance the probability to respond *rapace* /rapas/ “raptor” to *aigle* “eagle” because of the preceding rebus combination *rape* + *as* exerting a late priming effect (*rapace* “raptor”). To eliminate the possibility of this effect altogether, we created 7 supplementary rebuses (pairs of images) that we exclusively reserved for the control condition. The control condition was then constructed by randomizing target words over the 7 control rebuses.

In summary, we had a total of 14 target words, associated to 14 pairs of images composing their related rebuses; these are experimental stimuli. Each participant associated upon the 14 target words. 7 targets words were presented in the EX condition, preceded by their corresponding rebus images, and the remaining 7 target word were presented in the CR condition, preceded by rebus images coming from the supplementary pool of rebuses. For example, in the EX the target word (i.e., *félin* “feline”) is preceded by its associated rebus (in this case *panthère* /pāter/ “panther”) composed by the two images side by side (i.e., image of a *paon* /pā/ “peacock” and of *terre* /tɛr/ “earth”). In the CR the target word (i.e., *félin* “feline”) is preceded by an unrelated rebus (for example, *souplesse* /suples/ “flexibility”) composed by the two images side by side (i.e., image of a *soupe* /sup/ “soup” and of a *laisse* /les/ “leash”). Note that the *souplesse* rebus is thus one of the 7 supplementary French rebuses, designed specifically for the control condition, and therefore without associated target word (see section “2.1.2 Stimuli”).

Moreover, there were 2 variations for the presentation of our stimuli in Experiment 1: in the forward variation (FW) the images are presented in the forward order (i.e., the image of *paon*

/pā/ “peacock” followed by the image of *terre* /tɛr/ “earth”), and in the reverse variation (RV), in the reverse order (i.e., the image of *terre* /tɛr/ “earth” and of *paon* /pā/ “peacock”). Participants were assigned either to the FW variation or the RV variation of Experiment 1 and received the 14 target words—as said, 7 of them in the EX condition (preceded by their related rebus) and 7 in the CR condition (preceded by an unrelated rebus; see [Figure 3](#)). It is important to note that this procedure implies that participants never saw twice the same couple of images nor twice the same target word.

2.1.4. Funnel debriefing

To assess whether the participant stayed naive (or realized at one point) that the images presented side by side were forming a rebus, we composed an elaborate funnel debriefing, consisting of an 8-point evaluation. Note that the score is a *naivety score*: i.e., the higher the score, the naiver the participant. The structure of the debriefing is given in [Table 2](#). Debriefings are adapted according to the variations (FW or RV) but follow the same logical development.

As can be seen in [Table 2](#), the highest weight is given to the ability to give a concrete example (of a formerly presented or of another rebus) or to solve a presented rebus. However, even if the participant can formulate an approximation of the rebus principle without giving an example in questions 1 or 2, he is considered non-naive. The pivotal point of the debriefing is on question 4, there we give away the information that the images are word plays. If the participant solves the rebus (only) at this point, he is considered non-naive, except in the case he indicates that he only just now came to understand the principle thanks to the debriefing information. In other words, participants are considered naive if the naivety score is ≥ 5 on 8.

2.1.5. Procedure

The online survey started with written informed consent and inclusion criteria (older than 18 years and French speaking) followed by demographic and language questions (self-evaluation and FR12). Thereupon, participants received the following instruction (in French): “Please put yourself in the best possible conditions to respond (far from distractions). In the next task, what interests us is your spontaneous associations based on the word that is presented to you, that is, the first thing that comes to mind when you read this word. Even if your answers seem absurd or trivial, follow your intuition. There are no right or wrong answers. What you are asked to do is: 1 - look at the images on the screen for 4 seconds. 2 - read the word that will follow 3 - write the first six words that come to mind associated to that word.” The first stimulus appeared as two images side by side and stayed on screen for 4 seconds; then the stimulus as a target word appeared and participants received the following instruction: “Please write the first six words that come to mind upon the word:

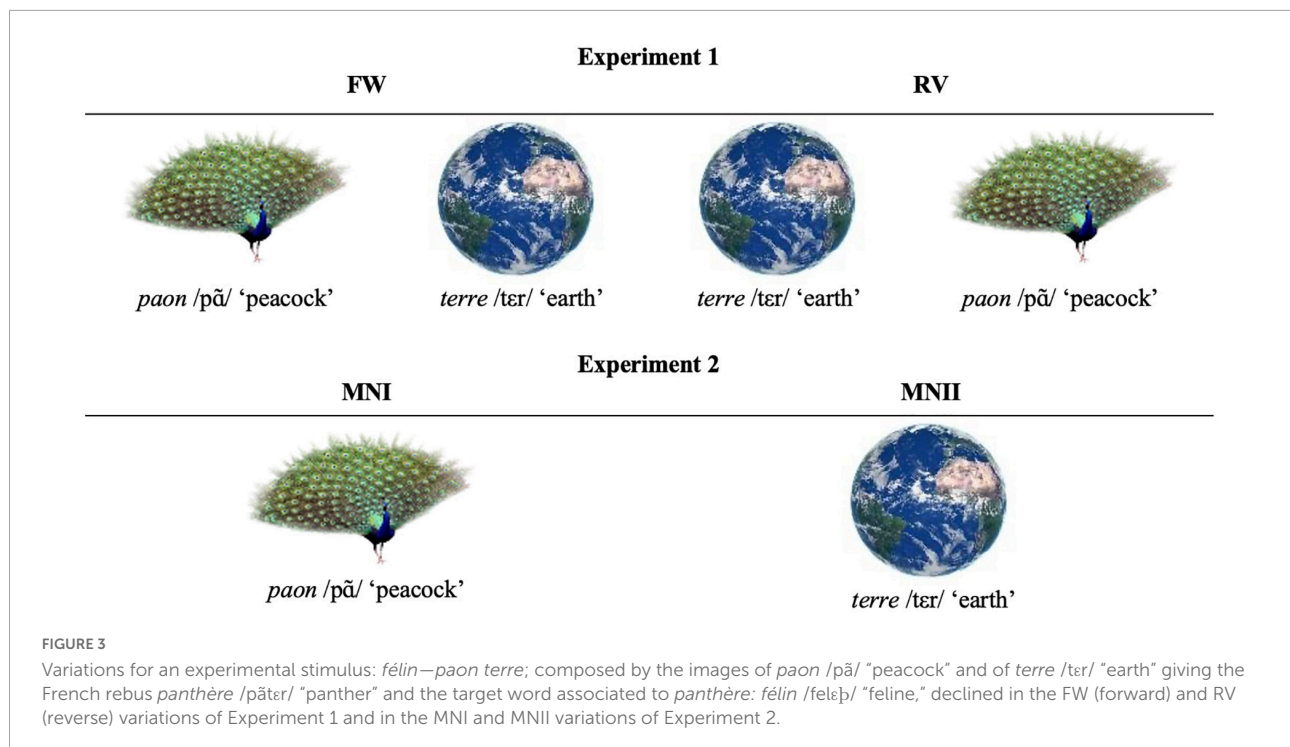


TABLE 2 Funnel debriefing in FW (forward) variation of Experiment 1.

Questions	Scores	
	<i>example</i>	<i>w/o example*</i>
1. Did you notice anything about the pairs of images that you have seen? If you can, give an example.	0	1 (correct elements)
2. Can you tell us something about these images? 	2	3 (correct elements)
3. The pairs of images you saw were <i>word plays</i> . If you can, give an example of such a word play. [<i>If an example of a correct rebus resolution is given:</i>] When did you understand this principle?***	4 (during the task) or 5 (after the task)	(go to 4)
4. The image pairs were word plays based on the names of the images. If you can, give an example.	6	(go to 5)
5. Take a good look at these two images. Can you solve this word play? 	7	8 (no correct answer all over)

For RV (reverse), all images were presented in reverse order. The rebus shown in question 2 is: *ver* /vɛr/ "worm," and *nid* /ni/ "nest," forming *verniss* /vɛrni/ "varnish"; the rebus shown in question 6 is: *poire* /pwar/ "pear," and *eau* /o/ "water" forming *poireau* /pwaro/ "leek." Note that as soon as either a correct rebus resolution or correct response elements are given, the scores are halted at the lowest naivety level.

*"example" means that participants receive the scores only if in the answer an example of a rebus resolution is given; "w/o example" means that participants receive the scores if correct elements of the rebus principle are given without giving an example.

***"during the task" was subdivided in: 1. since the first images seen; 2. during the task; 3. at the end of the task and 4. "after the task thanks to debriefing explanations." "After the task" here refers to 4. "after the task thanks to debriefing explanations" option only.

[target-word]." After the rebus task, participants answered the funnel debriefing.

2.1.6. Rebus scoring and analysis

We scored rebus resolution (RR) by attributing 1 point if the rebus resolution was found among the 6 written associations upon the related target word, independently of the fact if the rebus word appeared as the first or as a later association. This gives us a binary by-observation score (0 or 1). For example, for

the target word *félin* "feline" we counted if *panthère* "panther" occurred among the 6 associations to *félin* "feline" either in EX or CR of both FW and RV. We analyzed the RR score using a Generalized Linear Mixed Model computed by the *GAMLj* module (Gallucci, 2019) of Jamovi software (2022) (R Core Team, 2021). We used Generalized Linear Mixed Models to determine the impact of the EX and CR conditions and of the FW and RV variations on rebus resolution for each observation. We wanted to know if the rebus resolution score depends on

the target word being presented in the EX rather than in the CR condition and, simultaneously, if the order of the images (FW and RV variations) influences rebus resolution. We used the logit transformation for our analyses and included a by-subject and by-item intercept to consider differences among participants or items (in our case, target words), respectively. We added the naivety scores as well as the results on the FR12-test to control for the influence of naivety on rebus resolution, respectively of French language proficiency. We discarded model constructions with only either item or subject intercept and without Naivety Score or FR12 Score. In the results section, we present a model construction based on maximum likelihood and Akaike's Information Criterion (Akaike, 1974; please refer to [Supplementary material 3](#)). To show our results in bar-plots (see [Figure 4](#)), we used percentual rebus resolution (%RR), by dividing RR for a target word by the total number of times the target word appeared in each condition of each variation. This gives us a percentage of rebus resolution per target word. To analyze the %RR we used a Wilcoxon Rank Test. We performed Bayesian paired t-Test (Rouder et al., 2009; Eidswick, 2012) in case of non-significant results to support the absence of effect (null hypothesis). For the BF_{0+} we used a Cauchy prior width of 0.707. Bayesian factor BF_{0+} quantifies for one-sided null hypothesis (Bayesian paired t-Test made by the *Jasp* module on Jamovi; JASP Team, 2018; Morey and Rouder, 2018).

2.2. Results

2.2.1. Naivety results

On 788 participants that completed Experiment 1, 436 (i.e., 55.3 %) were naive to the rebus principle (see [Table 3](#) for demographic data).

There were no differences between the naive and non-naive participants on any of the demographic variables except for age. Naive participants are in average 2 years older than non-naive participants [$t(786) = 2.23$, $p = 0.01$, Cohen's $d = 0.16$]. Nonetheless rebus solving was not correlated to age ($r = -0.037$; $p = 0.29$; $N = 788$). As expected, the Naivety Score of the naive population is much higher (7 times higher) than the score of the non-naive participants. Accordingly, the mean of rebuses solved (the number of rebus solutions in the 6 written associations on the 7 rebuses) of the EX condition was higher in the non-naive population compared to the naive [means of 3.2 and 0.96; $t(786) = -20.2$, $p < 0.001$, Cohen's $d = -1.45$]. In line with our hypotheses, we are exclusively interested in unwitting rebus resolution; therefore, we exclude the non-naive participants in further analyses. Interestingly, the Naivety Score is also significantly and substantially higher in the FW as compared to the RV variation [resp. 6.6(1.0) and 7.5(0.8); $t(786) = 11.2$, $p < .001$, Cohen's $d = 0.823$].

2.2.2. Do people solve rebuses unwittingly?

We describe here the Model with four fixed effects: Condition, Variation, Naivety score and FR12 as well as subjects and item (target words) as random effects; our target is the RR. Results are presented in [Tables 4, 5](#) and [Figure 4](#).

The analysis of the fixed effects ([Table 4](#)) shows a significant effect of the condition EX-CR on RR score (estimate 0.18, $p = 0.02$). The variation (FW-RV) value is non-significant (estimate -0.127 , $p = 0.19$), the crossed effect between condition and variation is also non-significant (estimate -0.173 , $p = 0.28$). Finally, the effects the Naivety Score and of the French proficiency (FR12) are non-significant (Naivety Score: estimate 0.00, $p = 0.83$; FR12: estimate 0.05, $p = 0.14$). The analysis of deviance using Type III Wald chi-square tests shows that Condition (related rebus presentation in experimental condition

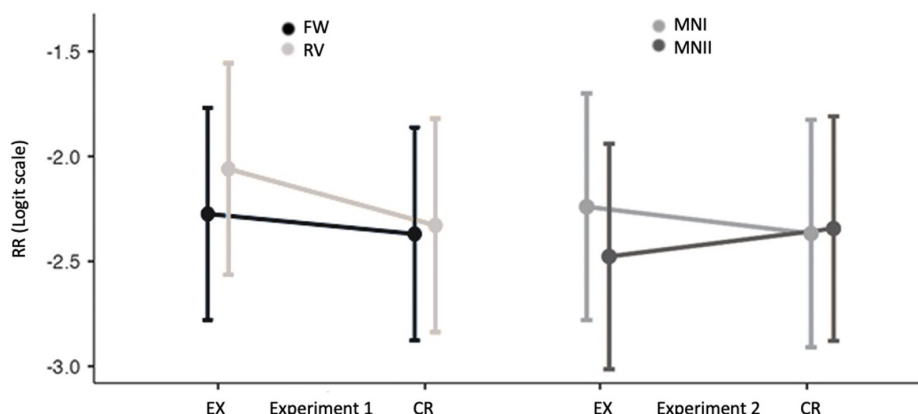


FIGURE 4

Fixed effects plots: linear predictor of the rebus resolution score (RR logit scale) for the condition experimental (EX) versus control (CR) and standard error for EX and CR conditions in the FW (forward) and RV (reverse) variations of Experiment 1 and EX and CR conditions in the MNI and MNII variations MNII (with only one of the two rebus images, respectively the first and the last) of Experiment 2.

TABLE 3 Mean (*M*) and standard deviation (*SD*) of the demographic data; *t*- and *p*-results on the independent sample T-test between naive and non-naive participants; effect-size for the significant effect; two-tailed *p*-value.

	Naive <i>N</i> = 436	Non-naive <i>N</i> = 352			
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>t</i>	<i>p</i>	Effect size
Naivety score (/8)	7.0(1.0)	0.9 (1.2)	76.5	<0.00**	5.48
Mean number of rebus solved (/7)	1.0 (0.9)	3.2 (2.0)	20.2	<0.00**	−1.45
Age (years)	32.1(13.0)	30.1(11.8)	2.23	0.02*	0.16
Education (years)	15.5(1.8)	15.4 (1.9)	0.16	0.8	
Self-assessment French proficiency (/7)	6.9 (0.1)	6.9 (0.1)	−0.6	0.4	
French proficiency test (/12)	11.0(1.2)	11.0 (1.0)	−0.5	0.6	

p*-value < 0.05; *p*-value < 0.001.

TABLE 4 Generalized linear mixed model for rebus resolution (RR) score in Experiment 1.

AIC	BIC	logLik		Deviance	def.resid	
4,202	4,256	−2,093		3,963	5,956	
	Fixed effects				Random effects	
					By subject	By item
Effect	Estimate	SE	<i>z</i>	<i>p</i>	SD	SD
(Intercept)	−2.188	0.415	−5.27	<0.001**	0.39	0.90
EX-CR	0.182	0.080	2.26	0.023*		
FW-RV	−0.127	0.097	−1.31	0.189		
EX-CR × FW-RV	−0.173	0.160	−1.07	0.281		
Naivety Score	0.009	0.047	−0.21	0.834		
FR12 Score	0.052	0.036	1.44	0.148		

Dependent variable is RR, the model family is binomial with a logit link. EX is the experimental condition; CR the control condition; FW the forward variation; RV the reverse variation, FR12 is the French proficiency test. Model formula: $\text{RebusScore} \sim 1 + \text{Condition} + \text{Variation} + \text{Naivety} + \text{FR}_{12} + \text{Condition:Variation} + (1 | \text{subjectID}) + (1 | \text{Target Word})$. **p*-value < 0.05; ***p*-value < 0.001.

or unrelated in control condition) is a significant predictor of Rebus Resolution score $\chi^2(\text{df} = 1) = 5.1408$, $p = 0.023$, confirming the effect of EX versus CR condition. Variation (forward versus reverse order of images presentation) is not a significant predictor [$\chi^2(\text{df} = 1) = 1.7220$, $p = 0.189$], nor is the interaction between Condition and Variation variables [$\chi^2(\text{df} = 1) = 1.1619$, $p = 0.281$]. Also, Naivety Score and French proficiency score (FR12) are not significant predictors of the RR score [resp., $\chi^2(\text{df} = 1) = 0.0440$, $p = 0.834$; $\chi^2(\text{df} = 1) = 2.0887$, $p = 0.148$]. In other words, naive people solve rebuses unwittingly and this rebus resolution is not affected by the image order, the degree of naivety, nor by the French proficiency.

We present in **Table 6** %RR in EX and CR for each target word. Note that the rebus is identified by its related target word, and (counterintuitively) not by its composing images as, in the CR-condition of the same rebus, the composing images originally belong to a different rebus and only the target word is kept constant (and therefore images and target word are unrelated). The high variability in rebus resolutions upon the target words in the EX condition is paralleled by a high

variability in rebus resolution in the CR condition, which is a good reassurance that the control condition indeed controls for the natural variability for the target words to spontaneously produce the rebus resolution in the EX condition. We present in **Figure 5** bar-plots associated with %RR.

Results of the Wilcoxon test showed non-significant difference between the FW and RV variations for the mean EX %RR (FW $M = 12.6$, $SD = 10.1$; RV: $M = 14.7$, $SD = 9.9$; $W = 33.0$, $p = 0.24$, $\text{BF}_{0+} = 1.8$) nor for the CR %RR (FW $M = 11.6$, $SD = 8.9$; RV: $M = 11.6$, $SD = 7.23$; $Z = 44.0$, $p = 0.63$, $\text{BF}_{0+} = 3.7$). Between FW and RV for both EX and CR conditions BF_{0+} qualifies for moderate evidence in favor of null hypothesis (see **Supplementary material 4** for the full detailed results). We further compared the %RR between EX and CR conditions and found, as expected, significant difference between EX and CR %RR in Experiment 1 (EX: $M = 13.7$, $SD = 9.7$; CR: $M = 11.6$, $SD = 7.9$; $Z = 81.0$, $p = 0.03$) with a medium effect-size ($r = 0.54$). In other words, there is inadvertent rebus priming by the visual environment of two images, independent of the order of these two images.

TABLE 5 Esteemed marginal means (*M*) and standard error (*SE*) for EX (experimental) and CR (control) rebus resolution (RR), as well as for EX and CR conditions in FW (forward) and RV (reverse) variations of Experiment 1.

Condition	Variation	<i>M</i>	<i>SE</i>
EX		0.1028	0.023
CR		0.0871	0.020
EX	FW	0.0933	0.021
CR	FW	0.0855	0.020
EX	RV	0.1131	0.025
CR	RV	0.0888	0.021

3. Experiment 2

3.1. Materials and methods

3.1.1. Participants

A total of 550 participants completed the entire online survey hosted by Limesurvey (all were majors and recruited on social media platforms). 80 participants were eliminated: 34 because they already have participated to one of our pilot studies and 46 because they were not native French speakers or had insufficient self-assessment scores for French language proficiency (see section “2.1.1 Participants” for further details). **Table 7** shows the demographic data for Experiment 2.

3.1.2. Stimuli, variations, and randomization

The same stimuli of Experiment 1 were proposed in Experiment 2 but in their “mono” version: with only one out of the two images composing the rebus, either the first one (i.e., the image of *paon* /pā/ “peacock”; MNI variation) or the second one (i.e., the image of *terre* /tɛr/ “earth”; MNII variation). In Experiment 2 we have followed strictly the same methodology as in Experiment 1: each participant was randomly assigned to either the MNI or the MNII variation and received 14 stimuli composed by one image and the target word (see **Figure 3**). The 14 stimuli were presented with 7 in the EX condition and 7 in the CR condition; for the CR condition, the 7 target words were randomly coupled with one image of a rebus coming from the supplementary control rebus pool, either the first image (MNI) or the last (MNII).

3.1.3. Funnel debriefing

To assess whether the participant stayed naive (or realized at one point) that the image presented might have influenced participants associations in the form of word plays we adapted the funnel debriefing of Experiment 1 for Experiment 2. The structure and the scoring principle were identical for the two experiments (see **Supplementary material 5** for the debriefing details of Experiment 2).

3.1.4. Procedure

The procedure of Experiment 2 is identical to the procedure of Experiment 1. The only difference is the instructions that were adapted since participants saw only one image (and not two as in Experiment 1).

3.2. Results

3.2.1. Naivety results

On 472 participants of Experiment 2, 469 (i.e., 99.3 %) were naive to the rebus principle (see **Table 8** for demographic data); only 3 were non-naive.

There were no significant demographic differences between naive and non-naive participant in Experiment 2. As expected, the Naivety Score of the naive population is much higher (more than 10 times higher) than the score of the non-naive participants. The mean number of rebuses solved by non-naive participants, resp. naive participants is 2.6 and 0.8 on a total of 7. The difference was significant [$t(470) = -3.56$, $p < 0.001$, Cohen's $d = -2.06$]. In line with our hypotheses, we are exclusively interested in unwitting rebus resolution; therefore, we exclude the non-naive participants in further analyses. The Naivety Score is essentially identical in the MNI and MNII variations of Experiment 2 [resp. 7.9(0.2) and 7.8(0.3); $t(470) = 0.44$, $p = 0.65$].

3.2.2. Inadvertent rebus priming by only one image?

We used the same model to analyze the unwitting effect of Condition (EX-CR) in the MNI and MNII variations of Experiment 2. Results are presented in **Tables 9, 10** and **Figure 4**.

The analysis of the fixed effects table (**Table 9**) shows non-significant effects of any of the variables of the model upon RR: Condition (estimate -0.00 , $p = 0.97$), Variation (estimate 0.10 , $p = 0.21$), interaction Condition*Variation (estimate 0.25 , $p = 0.10$), Naivety Score (estimate 0.26 , $p = 0.07$), the FR12 (estimate 0.04 , $p = 0.199$). The analysis of deviance using Type III Wald chi-square tests showed that none of the following are a significant predictor of Rebus Resolution score in Experiment 2: Condition $\chi^2(df = 1) = 0.0012$, $p = 0.971$, Variation $\chi^2(df = 1) = 1.5321$, $p = 0.216$, the interaction between Condition and Variation $\chi^2(df = 1) = 2.6704$, $p = 0.102$; Naivety Score $\chi^2(df = 1) = 3.2224$, $p = 0.073$; French proficiency score (FR12) $\chi^2(df = 1) = 1.6516$, $p = 0.199$. In other words, people do not solve rebuses unwittingly with only one image. We present the %RR in EX and CR for each target word in experiment 2 in **Supplementary material 6**. Even if we also calculate the %RR here (see section 2.1.6 “Rebus scoring and analysis”), participants in this Experiment 2 did not see the two images composing the rebus but only one, either the first (MNI) or the second (MNII). In other words, we did count e.g., *panthère* “panther” responses in Experiment 2 even if the

TABLE 6 Number of rebus resolutions, divided by the total number of times the target word appeared in each condition of each variation, presented as percentages (%RR), given by the naive participants to the target words in EX (experimental) and CR (control) conditions in, respectively, FW (forward) and RV (reverse) variations.

Target word	FW		RV	
	EX	CR	EX	CR
<i>Chine</i> —China	32.7	27.8	26.4	19.3
<i>Dakar</i> —Dakar	20.2	20.6	15.6	18.6
<i>danser</i> —dance	4.5	0.9	2.4	2.1
<i>félin</i> —feline	17.1	15.6	19.4	13.8
<i>fenêtre</i> —window	24.5	24.5	24.7	18.9
<i>feu</i> —fire	23.8	14.7	36.0	18.0
<i>fromage</i> —cheese	2.8	4.3	7.6	4.8
<i>jus</i> —juice	11.8	13.4	11.9	13.0
<i>ligne</i> —line	2.9	3.3	0.9	3.5
<i>montagne</i> —mountain	5.7	9.1	20.0	10.9
<i>nausée</i> —nausea	3.4	1.9	7.2	3.7
<i>toile</i> —canvas	19.1	18.0	16.0	23.2
<i>vaches</i> —cows	4.3	5.6	11.6	6.1
<i>voleurs</i> —thieves	2.7	2.7	5.8	5.2

Target words are presented in alphabetic order.

participant saw only the image of a “peacock” *paon* (MNI) or the image of the “earth” *terre* (MNII). We present in **Figure 5** bar-plots associated with %RR for Experiment 1 and Experiment 2. Results of the Wilcoxon test showed non-significant difference between the MNI and MNII variations within the EX (MNI $M = 13.9$, $SD = 9.9$; MNII: $M = 11.3$, $SD = 8.1$; $W = 67.0$, $p = 0.14$, $BF_{0+} = 1.5$) nor for the CR condition (MNI $M = 12.6$, $SD = 9.5$; MNII: $M = 12.5$, $SD = 10.4$; $Z = 55.0$, $p = 0.90$, $BF_{0+} = 3.7$). Finally, we found no significant differences in %RR between EX and CR conditions (EX: $M = 12.7$, $SD = 8.4$; CR: $M = 12.6$, $SD = 9.6$; $Z = 53.0$, $p = 0.64$; $BF_{0+} = 3.6$) within Experiment 2. Bayesian factor BF_{0+} which quantifies for one-sided null hypothesis (i.e., EX %RR is not larger than CR %RR) qualifies moderate evidence in favor of null hypothesis (there is 3.6 more chances that EX %RR is not larger than CR %RR; see **Supplementary material 7**). In other words, there is no significant evidence that inadvertent rebus priming occurs also by only one image.

4. General discussion

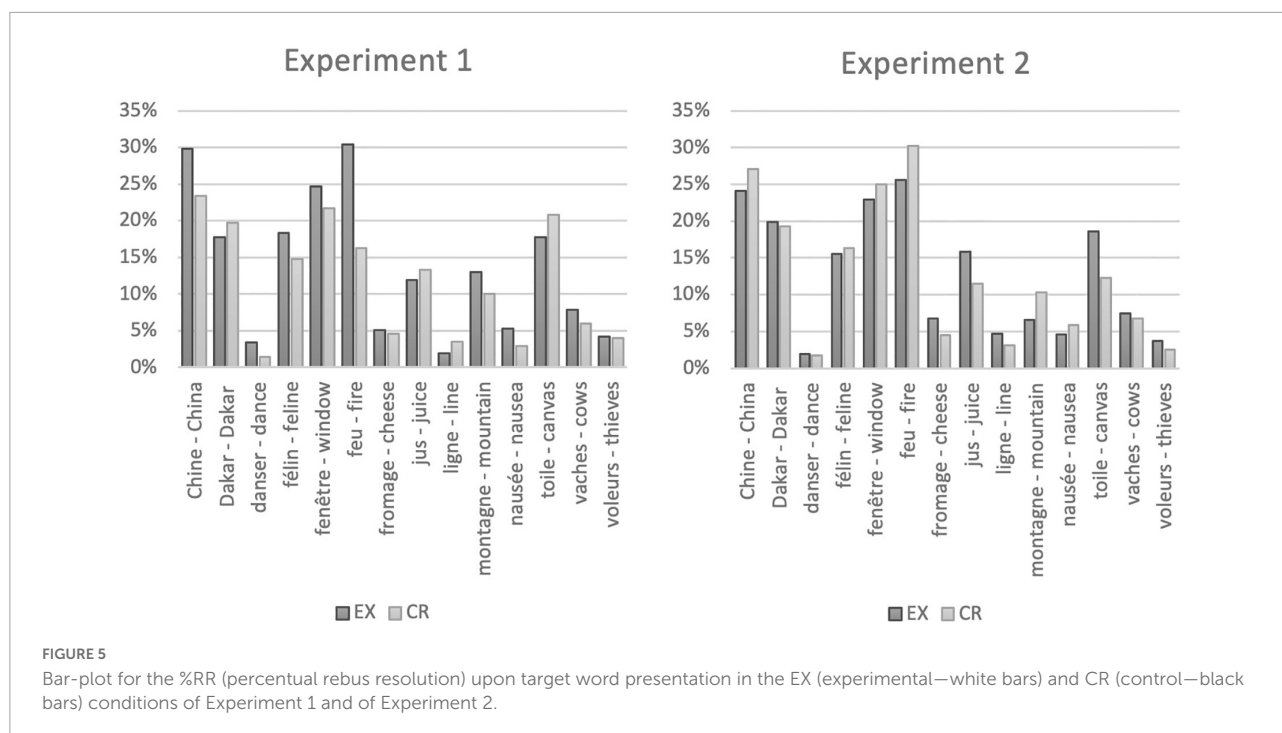
Our results show that people solve rebuses unwittingly, with images presented both in forward and reverse order (Experiment 1), and that this rebus resolution is not the result of phonological priming by just one of both images (Experiment 2). Therefore, the rebus resolution must be the result of the

simultaneous presentation of both images, and this in agreement with the fact that a rebus, indeed, implies the condensation of the two names of the images, following Freud's (1900, p. 295) definition of condensation: “Condensation is brought about [...] by latent elements which [are] fused into a single unity”. The rebus resolution is shown by our GLMM analysis which yields a significant effect of the difference between the experimental and the control conditions with an 0.182 estimate of the fixed coefficient ($p = 0.023$)². The rebus influence, though statistically significant, seems very modest when verified by occurrence, but we did expect as much beforehand. Also, Zwitterlood et al. (2018, p. 17), for example, comment “as often the case in word naming [in phonological-priming-by images], the effect was numerically small,” in line with the findings of the present study. Nevertheless, the results of the statistical model tell us that the images have a small, though significant, rebus effect upon ensuing verbal association events when considered in an event-by-event manner.

Remarkably, the rebus resolution works as well both in forward and in backward (reverse) presentation of the two images. We did expect this result as we proposed that rebus resolution is carried by the Freudian primary process, which works independently of spatial configuration (for which, as a matter of fact, spatial configuration has no significance; Freud, 1900, 1936; Lacan, 1964, p. 40). This is also a very “ecological” result, in other words, resembling real-life situations, as in daily life images appear in very chaotic, unpredictable spatial configurations, but our results confirm this should not impede their “rebus impact” on mental life.

As a matter of fact, even if the differences are not statistically significant, the absolute numbers are in favor of the backwards (reverse) as compared to the forward image presentation. The one big difference between both conditions is that significantly more people acquire insight (called “perspicacity” henceforth, as the opposite of “naivety”) in the rebus principle in the forward as compared to the backward condition during the presentations (55.6% vs. 26.3% non-naive participants in the total populations for FW, resp. RV variations; see “2.2.1 Naivety results”). This might lead to the suspicion that there was nevertheless more perspicacity in the FW condition than in the RV even after selection of the non-naive participants with our debriefing filter. However, it must be stressed that the naivety factor is not significant in our GLMM analysis, disproving any residual effect of the naivety score on rebus resolution. Even when we verify this influence more specifically for the rebus resolution in the FW condition, we obtain $r = 0.065$; $p = 0.34$ between experimental rebus resolution and naivety scores (now see **Supplementary material 8**) which shows that there is none such correlation, nor a positive one, nor—as might have been presumed here, a negative one. Therefore, if

² Effect sizes are not generally produced in generalized linear mixed models (McCullagh and Nelder, 1989).



we suspect a tendency of better rebus resolution in RV than in FW, this cannot be explained by a residue of conscious perspicacity with an inhibitory influence in FW, which might lead the participants to dismiss a seemingly irrational response at a conscious level: they would have insight but then would dismiss the rebus resolution, being blocked at the idea that e.g., a peacock and the earth are considered unrelated to a panther. However, in former research with the Shevrin lab we have found *unconscious inhibition* results: for example, in the so-called “pop look” studies of [Snodgrass and Shevrin \(2006\)](#), people who preferred to determine their identification response upon strictly subliminal stimuli by “looking” the best they could, gave *below-chance responses* as compared to people who claimed to be comfortable with letting responses spontaneously “pop up” (these latter people, gave above-chance correct identification responses). Similarly, in the [Villa et al. \(2006\)](#) palindrome study, supraliminal palindrome stimuli, such as e.g., “dog” (palindrome of “god”) did not prime “angel” responses when given a forced choice between 2 conscious targets. However, when the palindrome prime (e.g., “dog”) was presented strictly subliminally, non-anxious people gave *significantly below-chance* correct responses (here “angel”) while highly anxious people gave significantly above chance correct responses. In a final study with reversible primes (e.g., “door”; [Bazan et al., 2019](#)), highly defensive participants chose the phonological inverse (here, “road”) *significantly less than chance* in a forced similarity choice (against an unrelated target, e.g., “lung”), while lowly defensive participants chose the phonological inverse at above chance levels. These various results show that at a

subliminal, or unconscious level, inhibitory mechanisms start to come into play (see also [Shevrin, 1992](#)). It might be that this is also what we experience in the FW condition: one speculative explanation of our result tendencies might be that there is higher unconscious perspicacity in the FW than in the CR condition, in the same way that there is higher conscious perspicacity (filtered out by the debriefing). For a number of people this higher unconscious perspicacity might then induce unconscious dismissal of the rebus response (akin to a defensive or repressive response), explaining the response tendency in our research. However, this explanation remains at this stage speculative.

In the light of the former research on phonological priming by images ([Morsella and Miozzo, 2002](#); [Navarrete and Costa, 2005](#); [Roelofs, 2008](#); [Humphreys et al., 2010](#); [McQueen and Huettig, 2014](#); [Chabal and Marian, 2015](#); [Zwitserlood et al., 2018](#)), it is slightly suprising that we did not find significant rebus resolution when we only presented one

TABLE 7 Demographic data: Means (*M*) and standard deviations (*SD*) for *N* = 472 participants of Experiment 2.

	<i>M</i> (<i>SD</i>)
Age (years)	31.8 (12.8)
Education (years)	15.6(1.8)
Self-assessment French proficiency (/7)	6.9(0.1)
French proficiency test (/12)	11.1(1.1)
Gender (% women)	75.8
Handedness (% right-handed)	83.1

TABLE 8 Mean (M) and standard deviation (SD) of the demographic data for Experiment 2; *t*- and *p*-results for the independent sample T-test between naive and non-naive participants; two-tailed *p*-values.

	Naive N = 469	Non-naive N = 3			
	M(SD)	M(SD)	<i>t</i>	<i>p</i>	Effect size
Naivety score (on 8)	7.8(0.3)	0.6(0.6)	40.9	<0.00**	23.3
Mean number of rebus solved (/7)	0.8(0.8)	2.6(2.3)	3.5	<0.00**	−2.07
Age (years)	31.9(12.7)	23.6(3.2)	1.1	0.2	–
Education (years)	15.5(1.8)	15.6(1.1)	−0.0	0.9	–
Self-assessment French proficiency (/7)	6.9(0.1)	7.0(0.0)	−0.2	0.8	–
French proficiency test (/12)	11.1(1.1)	11.5(0.7)	−0.4	0.6	–

***p*-value < 0.001.

image (Experiment 2). However, it should be noted that our measures were quite different from these in the cited studies, namely spontaneous associations (as in contrast to naming tasks) and that moreover, we did not count the occurrence of the single names (e.g., the number of *dés* “dice” or *serre* “greenhouse” for the *dessert* rebus) or the number of their direct derivatives such as (e.g., *dessert*—*délice*) in the associations; both of these aspects reducing substantially our probability to find phonological priming effects of the images. In other words, Experiment 2 shows that only one image is not enough for rebus priming but did not show the absence of phonological priming by one image.

However, our principal result, showing inadvertent rebus priming by the visual environment, indeed confirms, with a different paradigm, former phonological-priming-by-images research. More specifically, Morsella and Miozzo (2002, p. 561) having indeed found that superposed images help, by mere phonology of their names (i.e., not through their semantic meaning) the naming of other images, in their discussion wonder “whether phonology is always activated for all the things that happen to fall on the perceptual system, or whether this unintentional activation of phonology occurs only in the context of speech tasks.” With our results, and considering the clinical data, we are inclined to confirm the first idea, the irrepressible activation for all things “that happen to fall on the perceptual system,” even when language is not specifically convened.

One element that remains unclear is the need for attention to the presented images. Indeed Roelofs (2003, p. 366) proposes that “attentional enhancements are a precondition for obtaining phonological activation from pictures.” In our research we have indeed asked to pay attention to the images, but we have not otherwise implied them in the ensuing experimental task (the free association upon the target words); in other set-ups participants were not specifically called to pay attention to Morsella and Miozzo (2002), Meyer and Damian (2007), or were explicitly invited to ignore (Humphreys et al., 2010) the relevant images—nevertheless, in all cases, attention was indeed involved, which goes in the direction of Roelofs’ proposition.

However, this question remains undecided as the phonological priming by images seems to work also in strictly subliminal conditions (Shevrin and Luborsky, 1961; Steinig et al., 2017). Moreover, a recent study of Chabal et al. (2022) shows that phonological priming of images occurs automatically and regardless of attention or memory demands.

Interestingly, Zwitterlood et al. (2018) show that the phonological-priming-by-images also works for action scenes, even when scenes do not have to be named—which widely opens up the naturalistic application of these findings to about anything present in the visual scene (and which is, by the way, also close to Freud’s idea, as his unresolved theoretical rebus example entails a *running* man, see above). This perspective upon larger implications as concerns the human mental world also resonates in Morsella and Miozzo’s (2002, p. 561) conclusions: “Beyond its implication for models of speech production, such a finding would bear on theories about the nature in which output programs are activated and selected in all actions, not just linguistic ones.” This track indeed leads to supposing structural linguistic influences not only in speech production, but “in all actions,” or even in all instances “in which output programs are activated,” i.e., more generally in the human behavioral and mental world—which would make sense with clinical observations.

We are hesitant to use the word “unconsciously” as our paradigm was a fully conscious, supraliminal paradigm and not a strictly subliminal paradigm, as was mobilized in the dream-rebus studies of Shevrin and Luborsky (1961) and Steinig et al. (2017). However, we had a strict debriefing procedure checking for the ability of the participant to report the word-play character of our images and selected only the naive participants in our analyses. As said, to completely exclude the possible contribution of any “perspicacity”-residues, which people wouldn’t have been able to report articulately, we also included “naivety” as a factor in our analyses and showed, by the absence of correlation that no rest of perspicacity influenced the results of our naive participants. In other words, we may be sure that the rebus resolution results found are

TABLE 9 Generalized linear mixed model for RR score in Experiment 2.

AIC	BIC	logLik	Deviance	def.resid	
4,365	4,419	−2,174	4,166	6,236	
Fixed effects				Random effects	
				By subject	By item
Effect	Estimate	SE	z	p	SD
(Intercept)	−4.876	1.232	−3.95	<0.001**	0.32
EX-CR	−0.002	0.078	−0.03	0.971	
MNI-MNII	0.105	0.085	1.23	0.216	
EX-CR * MNI-MNII	0.257	0.157	1.63	0.102	
Naivety Score	0.263	0.146	1.79	0.073	
FR12 Score	0.048	0.037	1.28	0.199	

Dependent variable is RR, the model family is binomial with a logit link. EX is the experimental condition; CR the control condition; MNI and MNII are the variations in Experiment 2 (with only one of the two rebus images, respectively, the first and the last), FR12 is the French proficiency test. Model formula: $\text{RebusScore} \sim 1 + \text{Condition} + \text{Variation} + \text{Naivety} + \text{FR_12} + \text{Condition:Variation} + (1 | \text{subjectID}) + (1 | \text{Target Word})$. ***p*-value < 0.001.

not the result of a conscious insight into our rebus paradigm. Thereby, people can resolve rebuses without being able to report doing so, this is, if nothing more, *unwittingly*. Moreover, without this being proved by our results, we suggest that the processes underlying this ability are the processes proposed by Freud for the functioning of the unconscious, namely primary process mentation, as these processes are supposed to be active continuously in the “background” (Freud, 1901, p. 239) of conscious processing. This would be coherent of the reported effects being independent of the image order. Thereby, we consider our supraliminal paradigm as being in fact a very ecological one, close to “normal” mental life.

Finally, to exclude any biases of French proficiency upon our results, we strictly selected participants with a very high self-assessment of their French proficiency. For an independent check for this influence, we added a quick self-designed French test (“FR12”), which turned up to confirm that in our select group of participants French proficiency was not a significant predictor of our rebus effect.

TABLE 10 Esteemed marginal means (*M*) and standard error (*SE*) for RR for EX (experimental) and CR (control) conditions in Experiment 2 and EX and CR conditions in MNI and MNII (with only one of the two rebus images, respectively the first and the last) variations of Experiment 2.

Condition	Variation	<i>M</i>	<i>SE</i>
EX		0.0857	0.020
CR		0.0860	0.020
EX	MNI	0.0954	0.023
CR	MNI	0.0850	0.021
EX	MNII	0.0770	0.019
CR	MNII	0.0869	0.021

Nevertheless, the ecological relevance of our findings might remain obscure as one might oppose: “We almost never encounter rebuses in real life”. Maybe there are more encounters than we think when a tie reaches a knee, and unconsciously hints to “tiny,” for example, but if we take the rebus-principle in a large, flexible way then we might find “ivan” in the divan or “Manhattan” when a man has his hat on, or we may feel primed to “arthrosis” when seeing a rose, to “carnival” upon seeing a car, to give a few possibilities. Opening this principle to action scenes, might imply that we are almost never unsolicited by slight phonological priming, as our attention is very often drawn to what is happening in the world. Even a marginal influence, integrated over so many visual takes a human being makes when being in the world, may create an accidental match between subject, situation, and environment, whereby the effect on human experience, is, we propose, a significant one. This is, it might help to explain why, for some subjects in specific situations, a black beetle elicits horror, the name of a shop induces well-being or the title of a book disgust, in a seemingly irrational, but nevertheless logically explainable way. Indeed, the relevance of these findings might be especially important in clinical situations: indeed, they might offer a rational ground for phenomena and their interpretation, which at first sight might seem completely “crazy”. For example, here is a remarkable excerpt of a patient with Tourette syndrome by the French neurologists Meige and Feindel (1902, p. 110) showing a conscious mobilization of the rebus principle: “to mimic the word commissary [*commissaire* in French] he [the patient] squeezed his right hand with his left hand [mimicking the comment] “how he squeezes!” [*comme il serre!* in French, phonologically close to *commissaire*] and to express doctor [*médecin* in French] he pretended to grasp on his chest imaginary breasts “my two breasts” [*mes deux seins* in French, phonologically close to *médecin*].”. Mostly the rebus structure of our “symptoms” remains obscure, even to ourselves, but

in some personality structures—such as psychotic structures or Tourette syndrome—the unconscious is thought to be at the surface and rebus reading might be mobilized consciously. Our results contribute to the idea that the rebus principle is a general principle for mental functioning, and thereby helps to understand phenomena described here.

Another element coming from a different field also hints on the significance of rebuses for the human mind: indeed, their attested existence traces back to the beginnings of civilization. In the first writing attempts in history (such as e.g., in Sumerian and in Egyptian hieroglyphs, around 3000–3100 BC) images were used to fix language on support. A real writing revolution occurred when these “images” were used, not only for the representation of things but also for the representation of sounds—i.e., when humans invented the rebus principle. The statue of Ramses II (dating around XIII century BC and conserved in the Museum of Egyptian Antiquities in Cairo), for example, shows a three part-composition depicting a solar disk, representing the sun God, called “Ra” in ancient Egyptian, a child called “mes” and a reed called “su” forming in fact Ram-ses-su or *Ramses*: this is, Ramses II’s name is on the statue in the form of a rebus, still accessible to deciphering after three thousand years. It is properly this *phoneme representation by images* called “phonogram” that in time leads to the inventions of alphabets and modern writing (Vernus, 2016, 2020). In other words, the rebus principle is a common principle to all writing systems and so much so that historians wonder if, to explain this universal phenomenon, we are not to suppose rebuses are a functioning principle of the human mind (Sington, 2020).

Given human’s apparently spontaneous appetite for and ease with such charades, it is slightly amazing that there aren’t more scientific studies on the topic. Among the few linguistic studies, most aren’t using our “picture + picture” principle but constitute miscellaneous combinations of words, letters, typographical writings, and pictures (see MacGregor and Cunningham, 2008; Salvi et al., 2015). As said, Freud is maybe among the first to give an explicit mental importance to rebuses, seeing them as the principle underlying dream formation, but also at the heart of “psychopathology of everyday life” such as in the forgetting of words (see the example of Signorelli, Freud, 1901). Therefore, it is in the psychoanalytic tradition, and specifically the tradition initiated by the American scientist and psychoanalyst, Howard Shevrin, that we find the most studies with direct relevance to the present one. Our research, even if it is directly inspired by Shevrin’s early rebus-dream studies, is, as said, however also different since, as a starter, it is a fully supraliminal paradigm. Moreover, Shevrin and colleagues did not use a priming set-up but asked for completely free associations (upon awakening participants in their REM-sleep). Therefore, given the discrepancies in experimental configurations, it is even more remarkable that the results are quite similar: people solve rebuses without knowing they do so, i.e., without any contribution of conscious cognitive

computing, i.e., in an unconscious (Shevrin and Luborsky, 1961; Steinig et al., 2017), or at least, in an unwitting way (our study). For an exercise which at first sight seems cognitively complex and even remains challenging when doing it fully conscious of the nature of the materials, this is astonishing and contributes to the idea that there is a complex mental life with highly developed capacities, which is at work without us being aware of its influence. This was also part of Jacques Lacan’s amazement in his famous “return to Freud” (Lacan, 1955). With the results of this strictly experimental research, it is also our endeavor to, again, encourage such a linguistic “return to Freud,” as we think the linguistic model is the best model to grasp amazing clinical phenomena as given above, and, more fundamentally to grasp the irreducibly irrational foundation of mental functioning.

Our research also ambitions to bring a nuance to the prevailing visual-semantic paradigm to understand the human grasping of the world (e.g., Kosslyn, 1994). Next to grasping visual images in the world and being potentially submitted to their associative semantic activation in our mind, we propose that there is also an unnoticed, i.e. very discrete, *formal* influence of the phonology of the subtexts of our experience—the names and phrases catching what we live—upon our mental processing, but which, integrated over so many experiences, has a significant role, and may help to explain what we tend to call “subjective,” i.e., our “unexplainable” tastes and dislikes, obsessions and fears, and sometimes our mental symptoms.

Even if this study was preceded by a considerable number of preliminary studies (15; $N = 3327$ in total), which helped us to perfect our methodological protocol and our stimuli, we still have several considerations, which are important to include for a further perfectioning in case of replication of the protocol. First, as concerns the target words, the best rebus resolution results were found with the target words with an associative strength to the rebus resolution of 0%, which means that participants are almost never spontaneously giving the rebus resolution as a first associate to the target word. Indeed, the resolution appears in further associations, and this situation is the one, we suggest, giving the best sensitivity for the difference between the experimental and the control set-up. We suppose that with target-rebus resolution couples at 0% associative strength between both, our results would have been even stronger. Second, as discussed, there was a considerable variability among the rebuses. Among the limits of our research, we did not take into account the information on word frequencies (composing words, target words, rebus resolution words) which might help to explain this variability. However, the item variability was taken into account in the GLMM analysis. Third, as the experimental set-up seems to be quite sensitive to any “hints” giving away the rebus-principle, we suggest having non-rebus filler presentations in a systematic way—this is, presenting two images not forming a rebus and with no obvious relationship to the target word—to “drown” the rebus presentations. In a final consideration, with

enough technical means, a *subliminal* (ideally, tachistoscopic) presentation of the rebus images, would of course be the litmus test to study unconscious processing.

In sum, our results with a supraliminal rebus paradigm show inadvertent rebus resolution of the visual environment as well as unwitting rebus resolution independently of image order, and show that this rebus solving is not due to the phonological priming by one of both images, but the result of a condensation of the names of both depicted objects. Our paradigm being an ecological one, this suggests that real life images appearing in chaotic spatial configurations have a discrete, formal influence upon our mental processing through the phonology of their related subtexts—the names and phrases associated to the experiences—which helps to explain subjective tastes and dislikes, and sometimes mental symptoms.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Comité d'Avis Ethique de la Faculté des Sciences Psychologiques et de l'Éducation - Advisory Ethics Committee of the Faculty of Psychological Sciences and Education. The participants provided their written informed consent to participate in this study.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication. Specifically, the experimental work was conducted by GO during the 6 years of her Ph.D. supervised by AB. This manuscript results from mutual support and the common work of authors that contributed equally to this article.

Funding

GO received a research grant from the Fund for Psychoanalytic Research (2017) of the American Psychoanalytic Association (APsaA - Fund for Psychoanalytic research) which paid for half a year of salary. We received functioning money from the International Psychoanalytic Association (IPA CERP

Grants n°4540) and from Mr. Robert Berry. For the open access publication fee, and we received the support from Mrs. Aliza Shevrin and Mr. Robert Berry.

Acknowledgments

GO's thesis started with 6 months grant from the American Psychoanalytic Association (APsaA) and continued with no funding for 5 years. We are heartfully thankful to the APsaA, Mr. Robert Berry, Mrs. Aliza Shevrin, and the International Psychoanalytic Association (IPA) for their respective modest contributions which for us turned out to be vital. GO also wants to thank the members of her Ph.D. doctoral committee: Stijn Vanheule at the University of Ghent and Regine Kolinsky and Nathalie Vecrussse at the Université libre de Bruxelles and Mr. Mathieu Frenot for his support during the data-analysis. This research has been done with the help of many Master students and trainees at the Université libre de Bruxelles and a special mention goes to Mario Bestion. We also want to thank the *Faculté des Sciences psychologiques et de l'éducation* of the *Université libre de Bruxelles* for the support received to finalize this publication. We thank our reviewers for the interesting comments that lead us to improve our paper, especially the reviewer that gave the beautiful example from Henry Meige's work. AB is thankful for the decisive mentoring received from Howard Shevrin in Ann Arbor (2003–2005) and beyond.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.965183/full#supplementary-material>

References

- Akaike, H. (1974). A new-look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 716–723.
- Bazan, A. (2007). *Des fantômes dans la voix. Une hypothèse neuropsychanalytique sur la structure de l'inconscient*. Montréal, QC: Liber.
- Bazan, A. (2011). Phantoms in the voice: A neuropsychanalytic hypothesis on the structure of the unconscious. *Neuropsychanal* 13, 161–176. doi: 10.1080/15294145.2011.10773672
- Bazan, A., Kushwaha, R., Winer, E. S., Snodgrass, J. M., Brakel, L. A. W., and Shevrin, H. (2019). Phonological ambiguity detection outside of consciousness and its defensive avoidance. *Front. Hum. Neurosci.* 13:77.
- Chabal, S., Hayakawa, S., and Marian, V. (2022). Language is activated by visual input regardless of memory demands or capacity. *Cognition* 222:104994.
- Chabal, S., and Marian, V. (2015). Speakers of different languages process the visual world differently. *J. Exp. Psychol. Gen.* 144, 539–550.
- Eidswick, J. (2012). A Bayesian alternative to null hypothesis significance testing. *Shinken Res. Bull.* 16, 2–15.
- Emmorey, K. (2002). “The effects of modality on spatial language: How signers and speakers talk about space,” in *Modality and structure in signed and spoken languages*, eds R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press), 405–421. doi: 10.1017/CBO9780511486777.019
- Freud, A. (1936). *The ego and the mechanisms of defence*. London: Karnac Books.
- Freud, S. (1897). *Letter from freud to fliess (1887–1994), december 29th, 1897. The complete letters of sigmund freud to wilhelm fliess*.
- Freud, S. (1895). “Project for a scientific psychology,” in *The standard edition of the complete psychological works of sigmund freud*, Vol. 1, ed. J. Strachey (London: Hogarth Press), 283–398.
- Freud, S. (1900). “The interpretation of dreams,” in *The standard edition of the complete psychological works of Sigmund Freud*, Vol. V, ed. J. Strachey (London: Hogarth Press), 339–751.
- Freud, S. (1901). “The psychopathology of everyday life,” in *The standard edition of the complete psychological works of Sigmund Freud*, Vol. VI, ed. J. Strachey (London: Hogarth Press), 1–279.
- Freud, S. (1905). “Jokes and their relation to the unconscious,” in *The standard edition of the complete psychological works of sigmund freud*, Vol. VIII, ed. J. Strachey (London: Hogarth Press), 9–236.
- Freud, S. (1909). “Notes upon a case of obsessional neurosis,” in *The standard edition of the complete psychological works of Sigmund Freud*, Vol. X, ed. J. Strachey (London: Hogarth Press), 153–318.
- Freud, S. (1915). “The unconscious,” in *The standard edition of the complete psychological works of Sigmund*, Vol. XIV, ed. J. Strachey (London: Hogarth Press), 159–215.
- Freud, S. (1927/1931). “The future of an illusion, civilization and its discontents and other works,” in *The standard edition of the complete psychological works of Sigmund Freud*, Vol. XXI, ed. J. Strachey (London: Hogarth Press).
- Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Available online at: <https://gamlj.github.io/>
- Holt, R. (1967). “The development of the primary process: A structural view,” in *Motives and thought: Essays in honor of David Rapaport*, ed. R. Holt (New York, NY: International Universities Press), 345–383.
- Huetting, F., and McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *J. Mem. Lang.* 57, 460–482. doi: 10.1016/j.jml.2007.02.001
- Humphreys, K. R., Boyd, C. H., and Watter, S. (2010). Phonological facilitation from pictures in a word association task: Evidence for routine cascaded processing in spoken word production. *Q. J. Exp. Psychol.* 63, 2289–2296. doi: 10.1080/17470218.2010.509802
- Jamovi (2022). *Jamovi – The Jamovi project. (Version 2.3) [Computer Software]*. Available online at: <https://www.jamovi.org>
- JASP Team (2018) *JASP. [Computer software]*. Available online at: <https://jasp-stats.org>
- Kapsambelis, V. (2005). L'humour des patients schizophrènes. *Rev. Psychother. Psychanalytique Groupe* 1, 63–71. doi: 10.3917/rppg.044.0063
- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/3653.001.0001
- Lacan, J. (1955). *Encore: Livre 20*. Paris: Editions du Seuil.
- Lacan, J. (1964). *Le séminaire, livre XI: Les Quatre concepts fondamentaux de la psychanalyse. Texte établi par Jacques-Alain Miller*. Paris: Editions du Seuil.
- Lacan, J. (1966). *Écrits*. Paris: Éditions du Seuil.
- Limesurvey. (2020). *GmbH. / LimeSurvey: An OpenSource survey tool / LimeSurvey GmbH*. Hamburg: Limesurvey.
- MacGregor, J. N., and Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behav. R. Methods* 40, 263–268. BRM.40.1.263 doi: 10.3758/BRM.40.1.263
- McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*, 2nd Edn. London: Chapman and Hall. doi: 10.1007/978-1-4899-3242-6
- McQueen, J. M., and Huetting, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Atten. Percept. Psychophys.* 76, 190–200. doi: 10.3758/s13414-013-0560-8
- Meige, H., and Feindel, E. (1902). *Les tics et leur traitement*. Paris: Masson et Cie.
- Meyer, A. S., and Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Mem. Cognit.* 35, 494–503. doi: 10.3758/BF03193289
- Meyer, D. E., Schvaneveldt, R. W., and Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Mem. Cognit.* 2, 309–321. doi: 10.3758/BF03209002
- Morey, R. D., and Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs. [R package]*. Available online at: <https://cran.r-project.org/package=BayesFactor>
- Morsella, E., and Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 555–563. doi: 10.1037/0278-7393.28.3.555
- Navarrete, E., and Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *J. Mem. Lang.* 53, 359–377. 2005.05.001 doi: 10.1016/j.jml.2005.05.001
- Olyff, G., Thieffry, L., and Bazan, A. (submitted). Normes associatives en français pour 243 mots dans une population adulte. *Eur. Rev. Appl. Psychol.*
- Préaud, M. (2004). Brève histoire du rebus français, suivie de quelques exemples de rebus pour la plupart inévidés. *Rev. Bibliothèque Nationale France* 18, 17–21.
- R Core Team. (2021). *R: A language and environment for statistical computing. (Version 4.1) [Computer software]*. Vienna: R Core Team.
- Rapaport, D. (1951). *Organization and pathology of thought; selected sources*. New York, NY: Columbia University Press. doi: 10.1037/10584-000
- Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the stroop task. *Psychol. Rev.* 110, 88–125. doi: 10.1037/0033-295X.110.1.88
- Roelofs, A. (2008). Tracing attention and the activation flow of spoken word planning using eye movements. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 353–368. doi: 10.1037/0278-7393.34.2.353
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237.
- Salvi, C., Costantini, G., Bricolo, E., Perugini, M., and Beeman, M. (2015). Validation of Italian rebus puzzles and compound remote associate problems. *Behav. Res.* 48, 664–685. doi: 10.3758/s13428-015-0597-9
- Shevrin, H. (1992). “Subliminal perception, memory, and consciousness: Cognitive and dynamic perspectives,” in *Proceedings of the adapted from papers presented at the american psychological association convention, Washington, DC, aug 22, 1986*, (New York, NY: Guilford Press).
- Shevrin, H., and Fritzler, D. E. (1968). Visual evoked response correlates of unconscious mental processes. *Science* 161, 295–298.
- Shevrin, H., and Luborsky, L. (1961). The rebus technique: A method for studying primary-process transformations of briefly exposed pictures. *J. Nerv. Ment. Dis.* 133, 479–488. doi: 10.1097/00005053-196112000-00002
- Shevrin, H., Bond, J. A., Brakel, L. A., Hertel, R. K., and Williams, W. J. (1996). *Conscious and unconscious processes: Psychodynamic cognitive, and neurophysiological convergences*. New York, NY: Guilford Press.
- Sington, D. (2020). *L'odyssée de l'écriture [documentary]*. *Arte France*. [23'23" and 49'55"]. Available online at: <https://www.arte.tv/fr/videos/083905-001-A/l-odysee-de-l-ecriture-1-3/>
- Snodgrass, M., and Shevrin, H. (2006). Unconscious inhibition and facilitation at the objective detection threshold: Replicable and qualitatively different

unconscious perceptual effects. *Cognition* 101, 43–79. doi: 10.1016/j.cognition.2005.06.006

Steinig, J., Bazan, A., Happe, S., Antonetti, S., and Shevrin, H. (2017). Processing of a subliminal rebus during sleep: Idiosyncratic primary versus secondary process associations upon awakening from REM- versus Non-REM-sleep. *Front. Psychol.* 8:1955. doi: 10.3389/fpsyg.2017.01955

Vernus, P. (2016). La naissance de l'écriture dans l'Égypte pharaonique: Une problématique revisitée. *Archéo Nil* 26, 105–134. doi: 10.3406/arnil.2016.1105

Vernus, P. (2020). Écriture, langue et images: La naissance des hiéroglyphes. *Claroscuro* 19, 1–23. doi: 10.35305/cl.vi19.51

Villa, K. K., Shevrin, H., Snodgrass, M., Bazan, A., and Brakel, L. (2006). Testing Freud's hypothesis that word forms and word meanings are functionally distinct in the unconscious: Subliminal primary process cognition and its links to personality. *Neuro Psychoan* 2, 117–138. doi: 10.1080/15294145.2006.10773521

Zwitserslood, P., Bölte, J., Hofmann, R., Meier, C. C., and Döbel, C. (2018). Seeing for speaking: Semantic and lexical information provided by briefly presented, naturalistic action scenes. *PLoS One* 13:e0194762. doi: 10.1371/journal.pone.0194762



OPEN ACCESS

EDITED BY
Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY
Roei Admon,
University of Haifa, Israel
Lisa Simon,
University of Haifa, Israel, in collaboration with
reviewer RA
Edward F. Pace-Schott,
Harvard Medical School, United States

*CORRESPONDENCE
Frans Schalkwijk
✉ f.schalkwijk@gmail.com

SPECIALTY SECTION
This article was submitted to
Psychodynamic Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 26 July 2022
ACCEPTED 26 January 2023
PUBLISHED 16 February 2023

CITATION
Schalkwijk F, Van Someren EJW, Nicolai NJ,
Uijtewaal JL and Wassing R (2023) From
childhood trauma to hyperarousal in adults:
The mediating effect of maladaptive shame
coping and insomnia.
Front. Hum. Neurosci. 17:990581.
doi: 10.3389/fnhum.2023.990581

COPYRIGHT
© 2023 Schalkwijk, Van Someren, Nicolai,
Uijtewaal and Wassing. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License (CC BY).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

From childhood trauma to hyperarousal in adults: The mediating effect of maladaptive shame coping and insomnia

Frans Schalkwijk^{1*}, Eus J. W. Van Someren^{2,3}, Nelleke J. Nicolai⁴,
Julia L. Uijtewaal¹ and Rick Wassing^{2,5}

¹Department of Forensic Special Education, University of Amsterdam, Amsterdam, Netherlands,
²Netherlands Institute for Neuroscience, Amsterdam, Netherlands, ³Department of Integrative
Neurophysiology, Center for Neurogenetics and Cognitive Research, Vrije Universiteit Amsterdam,
Amsterdam, Netherlands, ⁴Private Practice, Rotterdam, Netherlands, ⁵Woolcock Institute of Medical
Research, Sydney, NSW, Australia

Introduction: A new line of insomnia research focuses on the developmental trajectories from early life stress to insomnia in adulthood. Adverse childhood experiences (ACE's) might create a vulnerability for later maladaptive coping with distress, as seen in chronic hyperarousal or insomnia. In a functional magnetic resonance imaging (fMRI) study, failure to dissociate the neurobiological components of shame from autobiographical shameful memories in insomnia was reflected by continued activation of the dorsal anterior cingulate cortex (dACC), which may be a result of maladaptive coping in the wake of ACE's. Following up on that study, the current pilot study explores the relation between ACE's, shame coping-styles, adult insomnia, hyperarousal, and neurobiology of autobiographical memory.

Methods: We used existing data ($N = 57$) from individuals with insomnia ($N = 27$) and controls ($N = 30$), and asked these participants to complete the childhood trauma questionnaire (CTQ). Two structural equation models were used to test the hypotheses that shame-coping styles and insomnia symptom severity mediate the association between ACE's and (1) self-rated hyperarousal symptoms and (2) dACC activation to recall of autobiographical memories.

Results: For the association between ACE's and hyperarousal, there was a significant mediation of shame-coping style ($p < 0.05$). This model also indicated worse shame coping with more ACE's ($p < 0.05$) and worse insomnia symptoms with more ACE's ($p < 0.05$), but no association between shame coping and insomnia symptoms ($p = 0.154$). In contrast, dACC activation to recall of autobiographical memories could only be explained by its direct association with ACE's ($p < 0.05$), albeit that in this model more ACE's were also associated with worse insomnia symptoms.

Discussion: These findings could have an implication for the approach of treatment for insomnia. It could be focused more on trauma and emotional processing instead of conventional sleep interventions. Future studies are recommended to investigate the relationship mechanism between childhood trauma and insomnia, with additional factors of attachment styles, personality, and temperament.

KEYWORDS

insomnia, adverse childhood experiences (ACEs), childhood trauma, hyperarousal, dorsal anterior cingulate cortex (dACC), shame

1. Introduction

Adverse childhood experiences (ACEs) have a devastating effect on adult emotional functioning, especially if the ACEs are multiple, leading to higher risks for depressive illness, anxiety, addiction, suicidality, and medical illness (DeBellis, 2010). ACEs are negative experiences, such as emotional abuse or neglect, physical abuse or neglect, sexual abuse, and exposure to parental substance abuse, mental illness, or criminal behavior (Felitti et al., 1998). Other than traumatic situations, such as a car accident, a landslide, or war, ACEs are often inflicted by the caretakers, who in their position were expected to protect and regulate the emotions of the child. In psychoanalytic practice, we see patients with ACEs struggling with a chronic state of hyperarousal (Van der Kolk, 2014) and relationship anxiety (Morrison, 1989; Dorahy et al., 2016, 2017). Their self- and emotion-regulation strategies are limited, as they have internalized dysregulating caretakers that have been a threat to self-cohesion (DeYoung, 2015). Hyperarousal causes a collapse of reflective functioning; emotion and affects are no longer experienced as temporary (conscious and verbal) feeling states but as states reflecting identity (Fonagy et al., 2002). Thus, feelings, such as guilt and shame are experienced as “I am a bad person,” which reflects an equivalent mode of functioning (Allen et al., 2006).

Hyperarousal can be described not only by the above psychological factors but also by the physiological processes involved in arousal, including increased high-frequency electroencephalography (EEG) activity, hormone secretion (e.g., cortisol), limbic brain metabolism, and sympathetic activation (Riemann et al., 2010). Importantly, hyperarousal is considered one of the primary risk factors for the development of mental health problems.

In a series of papers reviewed below, we studied the relationship between hyperarousal and shameful memories from the past in patients suffering from insomnia. Following up on these papers, in the research presented here, we aim to further explain hyperarousal and shame coping by examining more severe negative experiences faced in childhood. We studied the mediating effect of shame coping and insomnia on ACEs and hyperarousal.

A first factor that links ACEs to hyperarousal may be maladaptive shame coping styles. Clinically, a distinction is made between adaptive and maladaptive shame. The former refers to bypassing feelings of shame, which contribute to adaptive self-esteem regulation and social communication (Schalkwijk et al., 2016). The latter is characterized by maladaptive thoughts or behaviors referring to negative self-esteem, and a desire to hide or escape, and is experienced as a negative feeling threatening identity (Luyten et al., 2002). ACEs negatively influence the development of a child's identity. Often, the sense of self is threatened and the child questions his or her role in these experiences, such as “Who am I, when they can do this to me?” (Lewis, 1971; Tracy et al., 2007; Lichtenberg et al., 2011; Schalkwijk, 2012). Usually, the personality of persons with many ACEs is characterized by feelings of shame, powerlessness, and lack of a sense of agency (Lanius et al., 2020). When circumstances are chronically unfavorable, as in the case of many ACEs, a child's personality is characterized by maladaptive shame-coping strategies. The shame manifests either as blame for their shortcomings (internalizing) or attributing blame to others (externalizing), which can lead to hyperarousal (Wassing et al., 2016) and mental health disorders (Campbell, 2016).

A second factor linking ACEs to hyperarousal is insomnia (Bader et al., 2007), which is a sleep disorder characterized by hyperarousal (Riemann et al., 2010) and is in itself a primary risk factor for the development of mental health problems (Manber et al., 2008; Baglioni et al., 2011; Christensen et al., 2016; Blom et al., 2017; Hertenstein et al., 2019). Bader et al. (2007) found that 46% of the participants with insomnia reported moderate to severe ACEs, and the number of ACEs has been found to be proportional to the severity of insomnia (Wang et al., 2016). Furthermore, Kajeepta et al. (2015) found evidence of an association between witnessing family conflict and sexual abuse, and insomnia. Moreover, more ACEs have been linked to disturbed rapid eye movement (REM) sleep in adulthood (Insana et al., 2012), which is a consistent marker not only of insomnia but also of other mental health disorders (Germain et al., 2008; Riemann et al., 2012; Pesonen et al., 2019).

Concerning these possible mediation effects between ACEs and hyperarousal, genome-wide association studies followed up by functional annotation analysis suggest strong correlations between the expression of insomnia risk genes, psychiatric traits, and the limbic circuitry (Jansen et al., 2019). Furthermore, a recent systematic review presented consistent findings of altered epigenetic modifications of the glucocorticoid gene NRC31 in adults exposed to ACEs (Lang et al., 2020), which is in line with persistent changes in the function of the hypothalamic-pituitary axis observed in mental health disorders (Griffiths and Hunter, 2014). Importantly, patients with insomnia do not necessarily have increased emotional reactivity to novel experiences, i.e., they do not show stronger initial emotional responses. A recent study with 25,758 participants of the United Kingdom Biobank showed no association between insomnia symptoms and limbic brain activation during an emotional task (Schiel et al., 2022). However, the differences from regular sleepers become evident after sleeping (Wassing and D’Rozario, 2022). Following a night with disturbed REM sleep, patients with insomnia show continued emotional reactivity, which may last for months or even years, and could explain a chronic hyperarousal state (Wassing et al., 2016). It has been proposed that adaptive emotion regulation depends on sound REM sleep (Walker, 2009; Walker and Van der Helm, 2009). During REM sleep, a selective memory consolidation process may support the disengagement of the emotional limbic and salience brain networks. Wassing et al. (2019a) found support for the hypothesis that maladaptive sleep in insomnia perturbs overnight emotion regulation and may contribute to chronic hyperarousal (Wassing et al., 2016). Wassing et al. (2019b) also found support for the hypothesis that chronic hyperarousal in insomnia disorder is associated with the failure to dissociate the limbic circuitry from long-term memory traces. Patients with insomnia who recalled autobiographical shameful memories showed stronger blood-oxygen-level-dependent (BOLD) responses in the dorsal anterior cingulate cortex (dACC), which is part of the limbic and salience circuitry. Recall of adverse emotional memories also activates the dACC in patients with post-traumatic stress disorder (PTSD; Whalley et al., 2013). PTSD can be conceptualized as a disorder of impaired fear-extinction memory, and delayed acquisition of the fear-extinction memory is linked with continued activation of the dACC (Seo et al., 2018). The dACC, otherwise known as the mid-cingulate cortex, has predominant connections with the amygdala, striatum, and dopaminergic and serotonergic systems, and is functionally linked with evaluating prediction errors, updating memory schema, and adapting/persisting behavioral responses (Kolling et al., 2016; van Heukelum et al., 2020).

Considering these findings, we aimed to identify the dACC as the neurobiological expression of hyperarousal. Since the dACC is most consistently shown to be involved in emotional memory tasks in insomnia and PTSD populations (Whalley et al., 2013; Seo et al., 2018; Wassing et al., 2019b) we limited our analyses to the dACC and not to other nodes of the salience or limbic network.

Following this line of research, in this exploratory study, we asked ourselves where the vulnerability for hyperarousal and aberrant neurobiology of autobiographical memory in people with insomnia might originate. Could ACEs lead to the vulnerability of insomnia by weakening the stress regulatory processes?

To the best of our knowledge, this is the first exploratory study that investigated the relationship between ACEs, maladaptive shame coping, insomnia, hyperarousal, and dACC activation in the recall of autobiographical memories. In order to evaluate these relationships, we employed structural equation modeling with one direct and two parallel mediation paths so that all covariance between the variables of the hypothesized paths could be fitted to the observed data at once. We used two structural equation models to test the hypotheses that shame-coping styles and insomnia symptom severity mediate the association between ACEs and (1) self-rated hyperarousal symptoms, and (2) dACC activation in recalling autobiographical memories. Furthermore, we expected that more ACEs would result in worse maladaptive shame-coping styles and insomnia severity.

2. Materials and methods

2.1. Participants

The 57 participants in this study were described by Wassing et al. (2019b). These participants included 27 patients that fulfilled the diagnostic criteria for insomnia disorder [mean, standard deviation (SD) age, 45.5 (13.4) years; 10 men; ISI range 8–26] and regular sleepers [$n = 30$; mean [SD] age, 42.4 (15.8) years; 17 men, ISI range 0–14]. The participants answered the self-report questionnaires approximately 1 week before the functional magnetic resonance imaging (fMRI) scan.

Since the childhood trauma Questionnaire (CTQ) (Bernstein et al., 2003) was not part of this dataset established in 2015–2016, in February 2021 we asked these participants to complete the CTQ, and 44 participants responded (77%, $n = 23$ patients with insomnia, $n = 21$ regular sleepers).

This study was approved by the Ethical Committee of the Department of Education of the University of Amsterdam (approval number 3421). Written informed consent was obtained from the participants prior to enrollment in the study.

2.2. Measures

2.2.1. ACEs

Adverse childhood experiences were assessed using the Dutch version of the CTQ (Bernstein et al., 1997; Thombs et al., 2009; Spinhoven et al., 2014). It consists of 25 items rated on a five-point Likert scale (1 = never true, 5 = very often true) and measures five categories of trauma: emotional abuse, physical abuse, sexual abuse, physical neglect, and emotional neglect. An example of a physical neglect item is “During my childhood. . . I had to wear dirty clothes.”

The CTQ demonstrated a Cronbach's alpha of 0.95 for the total scale. The Cronbach's alpha for the subscales were 0.91 for physical abuse, 0.89 for emotional abuse, 0.95 for sexual abuse, 0.63 for physical neglect, and 0.91 for emotional neglect.

2.2.2. Coping with shame

Coping with shame was assessed using the Compass of Shame Scale (CoSS) (Elison et al., 2006; Schalkwijk et al., 2016). This self-report questionnaire consists of 15 shame-inducing situations and six answer categories. The participants were asked about their reaction to a situation, such as “When an activity makes me feel like my strength or skill is inferior,” and a reaction, such as “I get mad at myself for not being good enough,” on a five-point Likert scale (1 = never, 5 = almost always). This scale measures two maladaptive shame-coping styles: internalizing and externalizing. Internalizing shame coping contains answers in which the person blames himself or herself for being ashamed (“I brood over my flaws”) or the person avoids the shame-provoking situation (“I want to disappear”). Externalizing shame coping includes items in which the person blames others for feeling shame (“I get irritated with other people”) or denies that the situation evokes shame (“I act as if it isn't so”). Two other categories assess adaptive shame coping (“I think, that was stupid of me, but that's what happens to everyone”) and proneness to experience shame (“I am ashamed”). The Cronbach's alpha for the subscales were as follows: avoidance, 0.75; attack other, 0.76; attack self, 0.86; withdrawal, 0.75; adaptive, 0.78; and shame proneness, 0.87.

2.2.3. Insomnia

Insomnia symptom severity was assessed using the Insomnia Severity Index (ISI), a self-report measure used to assess the nature, severity, and impact of insomnia (Bastien et al., 2001). A five-point Likert scale was used (0 = no problem, 4 = severe problem), with a total score ranging from 0 to 28. An example item is “Please rate the current SEVERITY of your insomnia problems: difficulty falling asleep.” The ISI demonstrated a Cronbach's alpha of 0.76.

2.2.4. Hyperarousal

Hyperarousal was assessed using the Hyperarousal Scale (HS; Pavlova et al., 2001), a self-report questionnaire consisting of 26 items in the form of statements. Participants judged the statements using four different responses (0 = not at all, 1 = a little, 2 = quite a bit, and 3 = extremely). An example item is “I keep thinking about things long after they happened.” The HS demonstrated a Cronbach's alpha of 0.81 (Bruno et al., 2020). Note that two out of the 26 items overlap with those in the ISI: “I am slow to awaken in the mornings” (HS) may be anticorrelated with “problems waking up too early” (ISI), and “I have trouble falling asleep” (HS) is likely to be correlated with “difficulty falling asleep” (ISI).

2.2.5. dACC reactivity to recall of autobiographical memories

The dACC BOLD responses to shameful experiences were obtained during fMRI recordings. For a detailed description of the procedures, functional and structural MRI acquisition, data processing, and statistical analysis see Wassing et al. (2019b). In summary, 1 week prior to the fMRI scan, participants were asked to think of up to ten shameful past experiences, and for each to write down four cue words that, when prompted, allowed them to recall and relive the experience. In addition, for each

shameful experience, participants were instructed to also write four cue words for neutral events from the same period (e.g., traveling to work/school). During the fMRI scan, the cue words of each of the experiences were presented on the screen for 16 s. A total of 10 blocks of shameful conditions and 10 blocks of neutral conditions were presented in pseudorandom order. The raw fMRI data were preprocessed using FMRIB Software Library (FSL)¹ including corrections for inhomogeneities in the B0-field, motion correction using registration and independent component analysis with automatic removal of motion artifacts (ICA-AROMA) (Pruim et al., 2015), spatial smoothing (5 mm full width at half maximum kernel), grand-mean intensity normalization, nuisance signal regression (from cerebrospinal fluid and white-matter brain masks), and temporal high-pass filtering (1/90 Hz). The mean dACC activation for each participant was obtained using linear regression models in FSL FMRI Expert Analysis Tool (FEAT) (Woolrich et al., 2001) including a contrast on the regressors (shameful vs. neutral) and a mask of the dACC obtained in Wassing et al. (2019b).

2.3. Statistical approach

The hypotheses were tested using structural equation modeling in R (version 4.1.2, R Foundation, Vienna, Austria) and RStudio (2022.02.3 Build 492) using the *lavaan* package (R Core Team, 2021). Structural equation modeling is a statistical approach that combines confirmatory factor analysis to extract latent variables from observed variables with regression analysis to estimate and evaluate the statistical significance of hypothesized paths. Instead of performing separate linear models per path, the technique uses the multivariate covariance structure of the observed data to fit the model all at once. We employed two models, one with self-rated hyperarousal as the outcome measure, and one for dACC reactivity to recall autobiographical memories. Path coefficients were obtained using maximum likelihood estimators and considered significant at a two-sided alpha threshold of 0.05. Finally, full mediation was indicated by a significant mediation path estimate in combination with a non-significant direct path estimate between ACEs and the outcome variable. A partial mediation was indicated if both the direct path estimate and the mediation path estimate were significant.

2.3.1. ACEs and self-rated hyperarousal

The measurement model included four latent variables: (1) “ACEs” were measured by the observed scores on the CTQ subscales of emotional neglect, physical neglect, emotional abuse, physical abuse, and sexual abuse; (2) “shame” was measured by the CoSS subscales of shame-proneness, adaptive shame coping, and internalizing and externalizing maladaptive coping; (3) “insomnia” was directly measured by the ISI total score; and (4) “hyperarousal” was directly measured by the hyperarousal-scale total score. The regression model included three equations, all of which included age and sex as confounders. First, shame was predicted by ACEs. Second, insomnia was predicted by ACEs and shame. Third, hyperarousal was predicted by ACEs, shame, and insomnia scores. No moderating effects were considered. Mediation path estimates were computed

where shame, insomnia, and both shame and insomnia mediated the association between ACEs and hyperarousal.

2.3.2. ACEs and dACC reactivity to recall autobiographical memories

This structural equation model was similar to the one described in the previous subsection, except that the “hyperarousal” latent variable was replaced with the dACC BOLD responses for relived shameful experiences. Note that the path estimates in the two models between the same variables (e.g., ACEs to insomnia) could be different because the variance explained in ACEs by hyperarousal (model 1) was different from its variance explained by dACC BOLD responses (model 2). This means that the remaining variance in ACEs that was *not* explained by the outcome variable but *was* explained by insomnia could be different too.

3. Results

3.1. Descriptive ACEs statistics

The mean and SD values of the five CTQ subscales are shown in Table 1. Each subscale contained five question items, and their scores ranged from 5 to 25. Sexual abuse had the lowest mean score ($M = 5.27$), close to the lowest possible score. Emotional neglect had the highest mean score ($M = 12.16$), which was half of the highest possible subscale score.

3.2. Correlation structure among observed variables

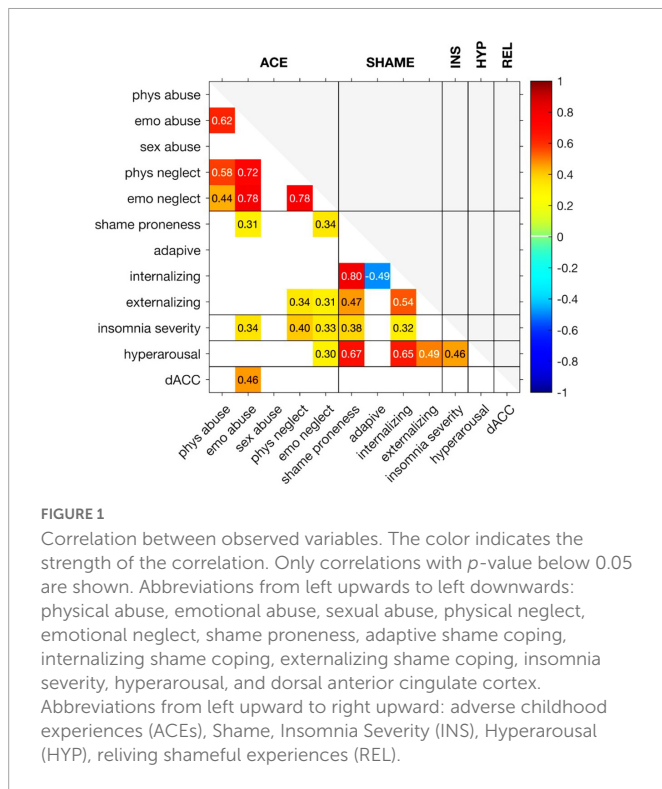
The correlations between observed variables are shown in Figure 1. Concerning the CTQ subscales, the ISI total score was correlated with emotional abuse ($r = 0.34, p < 0.03$), physical neglect ($r = 0.40, p < 0.007$), and emotional neglect ($r = 0.33, p < 0.03$). Emotional abuse was also correlated with shame proneness ($r = 0.31, p < 0.04$), physical neglect with maladaptive externalizing shame coping ($r = 0.34, p < 0.03$), and emotional neglect was correlated with both shame proneness ($r = 0.34, p < 0.03$) and maladaptive externalizing shame coping ($r = 0.31, p < 0.05$). Most CTQ subscales were not associated with hyperarousal or dACC responses to reliving shameful experiences. Only emotional neglect correlated significantly with hyperarousal ($r = 0.30, p < 0.05$), and emotional abuse with dACC responses to reliving shameful experiences ($r = 0.46, p < 0.002$).

Concerning the CoSS subscales, shame proneness was correlated with the ISI total score ($r = 0.38, p < 0.02$) and hyperarousal ($r = 0.67$,

TABLE 1 Descriptive statistics of adverse childhood experiences (ACEs).

	Mean	SD	Range
Emotional neglect	12.16	5.55	5–25
Physical neglect	7.05	2.16	5–12
Emotional abuse	8.68	4.20	5–20
Physical abuse	6.33	3.04	5–20
Sexual abuse	5.27	0.90	5–10

¹ <https://fsl.fmrib.ox.ac.uk>



$p < 10^{-6}$). In addition, maladaptive internalizing shame coping was correlated with the ISI total score ($r = 0.32$, $p < 0.04$) and hyperarousal ($r = 0.65$, $p < 10^{-5}$). Finally, maladaptive externalizing shame coping was correlated with hyperarousal ($r = 0.49$, $p < 0.001$).

The ISI total score was correlated with hyperarousal ($r = 0.46$, $p < 0.002$).

3.3. Structural equation modeling

3.3.1. ACEs and hyperarousal

All the model-fit indices combined indicated a satisfactory to a good fit. The Root Mean Square Error of Approximation (RMSEA) was 0.09 ($p = 0.14$; the null hypothesis that the RMSEA was less than or equal to 0.05, could not be rejected, indicating a good fit). However, the discrepancy between the observed and fitted covariance matrices was barely significant ($\chi^2(56) = 74.7$, $p = 0.048$). The sample-size insensitive metrics also indicated a satisfactory to a good fit, that is, the comparative fit index = 0.92 (good) and the Tucker–Lewis index = 0.89 (satisfactory > 0.90). The factor loadings of the observed variables on the latent variables are shown in **Figure 2**. Of note, sexual abuse was loaded onto the ACEs latent variable with trend-level significance ($p = 0.09$). The structural equation model for hyperarousal is shown in **Figure 2**.

3.3.2. Path estimates

Adverse childhood experiences did not directly predict hyperarousal ($p = 0.98$), but there were significant associations between ACEs and shame [β (SD) = 0.05 (0.02), standardized $\beta = 0.33$, $Z = 2.08$, $p < 0.04$] and ACEs and insomnia [β (SD) = 0.58 (0.24), standardized $\beta = 0.36$, $Z = 2.42$, $p < 0.02$]. However, insomnia was not predicted by shame ($p = 0.17$). Hyperarousal was predicted by shame [β (SD) = 7.30 (1.42), standardized $\beta = 0.65$, $Z = 5.15$, $p < 10^{-6}$] and insomnia [β (SD) = 0.21 (0.11), standardized $\beta = 0.22$,

$Z = 1.96$, $p < 0.05$]. The combined path estimate from ACEs *via* insomnia to hyperarousal was not significant [β (SD) = 0.12 (0.08), standardized $\beta = 0.08$, $Z = 1.50$, $p = 0.13$]. Finally, the combined path estimate from ACEs *via* shame to hyperarousal was significant [β (SD) = 0.33 (0.17), standardized $\beta = 0.21$, $Z = 1.98$, $p < 0.05$], indicating that the association between ACEs and hyperarousal was fully mediated by shame.

3.4. ACEs and dACC reactivity to recall of autobiographical memories

All model-fit indices combined indicated a good fit. The RMSEA was 0.06 ($p = 0.39$), and there was no significant discrepancy between the observed and fitted covariance matrices ($\chi^2(56) = 64.9$, $p = 0.20$). The comparative fit index (0.95) and Tucker–Lewis index (0.94) also indicated a good fit.

3.4.1. Path estimates

Adverse childhood experiences directly predicted the dACC responses to relived experiences [β (SD) = 0.07 (0.03), standardized $\beta = 0.39$, $Z = 2.38$, $p < 0.02$]. Furthermore, while ACEs did not predict shame in this model ($p = 0.12$), ACEs did predict insomnia [β (SD) = 0.62 (0.24), standardized $\beta = 0.38$, $Z = 2.62$, $p < 0.01$], although insomnia was not predicted by shame ($p = 0.15$). Finally, there were no associations between dACC responses to relived experiences and shame ($p = 0.85$) or insomnia ($p = 0.82$), which also indicated that no mediation paths were significant ($p > 0.82$). The structural equation model for dACC reactivity to recall autobiographical memories is shown in **Figure 3**.

4. Discussion

This exploratory study investigated the mediating effect of dysfunctional shame regulation strategies and insomnia between ACEs and emotionally experienced chronic hyperarousal, and its potential neurobiological underpinnings.

Our main interest in this study was the mediating effect of dysfunctional shame regulation strategies and insomnia between ACEs and emotionally experienced chronic hyperarousal and dACC responses to reliving shameful experiences. The bivariate correlation structure showed that, from the five groups of ACEs, emotional neglect seemed to have the most impact on hyperarousal and emotional abuse on dACC responses to reliving shameful experiences. Indeed, emotional abuse and neglect have been linked with the most long-term detrimental effects (Strathearn et al., 2020). Clinically speaking, maladaptive shame regulation strategies have been developed. Note that in this first structural equation model, insomnia did not significantly mediate the association between ACEs and hyperarousal, whereas indeed a significant mediation effect was found for maladaptive shame coping styles.

In the second structural equation model, our interest was the mediating effect of dysfunctional shame regulation strategies and insomnia between ACEs and the dACC response to reliving shameful memories. In contrast to the first model, we found no mediational effects for maladaptive shame regulation, nor for insomnia symptom severity. In a study by Wassing et al. (2019b), the insomnia patients, but not the controls, demonstrated dACC responses to the shame-inducing recall of autobiographical memories. In the

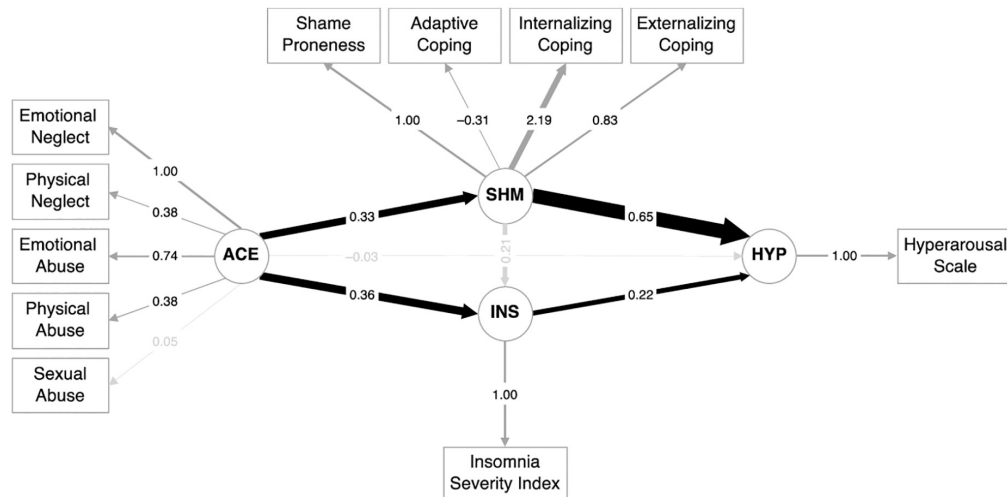


FIGURE 2
Structural equation model for hyperarousal.

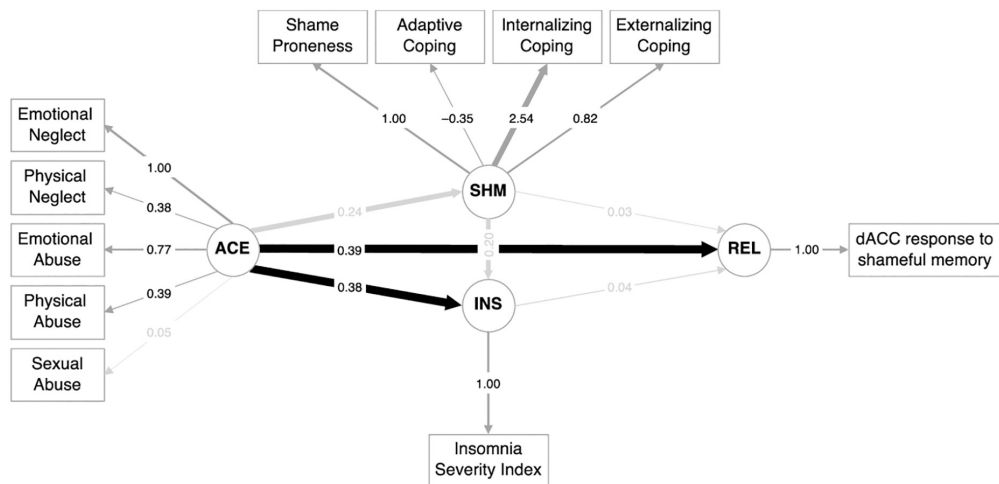


FIGURE 3
Structural equation model for dorsal anterior cingulate cortex (dACC) reactivity to recall of autobiographical memories.

current study, we now show that more ACEs relate to more severe insomnia symptoms and that insomnia severity did not mediate the association between ACEs and dACC responses. This may indicate that ACEs independently contribute to dACC responses when recalling shameful experiences. According to [Van der Kolk's \(2014\)](#) book title, the body does keep score of these events leading to hyperarousal. Our findings support the importance of explicitly addressing the component of shame in ACEs during treatment.

Secondly, we discuss the bivariate correlation structure reported in [Figure 1](#). Here we analyzed the premise that more ACEs lead to increased insomnia ([Bader et al., 2007; Wang et al., 2016; Muscatello et al., 2020; Reffi et al., 2022](#)). This was confirmed: having experienced more ACEs was predictive of insomnia severity in adults. However, this was found merely for the ACEs of emotional neglect, emotional abuse, and physical neglect. Differently from what happens with ACEs that include physical intrusions, such as sexual or physical abuse, other types of ACEs, such as emotional neglect, emotional abuse, and physical neglect, may go unnoticed as a child may never address or voice these experiences. In these cases, children tend to react physically and emotionally to these experiences with

maladaptive coping strategies. Prior work has shown that emotional abuse and neglect can have the most detrimental effects long term ([DeBellis, 2010](#)). These kinds of ACEs are associated with much more adverse outcomes in areas of cognition and education, psychological and mental health, addiction and substance use, and sexual and physical health than the physically intrusive ACEs ([Strathearn et al., 2020](#)). In our study, we showed the association between maladaptive shame regulation strategies and experiencing a subjective state of chronic hyperarousal ([Schore, 2003; Porges, 2011](#)). Sexual and physical abuse, however, is associated with the violation of body integrity, which is a far more obvious violation for a child to notice and these might lead to other coping strategies. A typical outcome of sexual abuse is lifelong PTSD and depression ([Strathearn et al., 2020](#)), dissociation ([Draijer and Langeland, 1999; Dorahy et al., 2017; Vonderlin et al., 2018; Nicolai, 2020; Voestermans et al., 2021](#)), and self-destructive behavior. Dissociation protects the individual from experiencing the painful emotions experienced during the abuse, such as physical pain, shame, disgust, and helplessness. Also, there is no sexual abuse without neglect ([Draijer, 1990](#)). In other words, ACEs seldom take place alone.

This pilot study has several limitations. The number of participants was small and they were recruited based on insomnia rather than trauma. Therefore, those who were physically and sexually abused probably were underrepresented in our sample and the distribution of male and female participants was uneven. Another limitation was caused by the psychological effect of chronic arousal mentioned in the introduction. We have discussed that chronic hyperarousal hampers reflective functioning and people have limited capacity for recognizing (conscious and verbal) feeling states. Feelings cannot be differentiated or reflected upon, so feelings of shame, guilt, and anxiety end up confirming the cognition of preconceived notions, such as that the individual who suffered ACEs is bad, stupid, and disgusting. The found effect of ACEs on chronic hyperarousal might thus be underreported, as people have difficulties in reflecting on their (conscious and verbal) feeling states.

A few recommendations are suggested from this study. A larger sample from a variety of clinical institutions would better represent a population with insomnia. In addition, future research should consider that adult survivors of childhood abuse have different trajectories leading to insomnia (Steine et al., 2019). For example, another interesting mediating variable might be attachment style, as research has shown that shame and hyperarousal are associated implicitly with attachment styles, the way people engage in close and intimate relationships (Van der Kolk, 2014). Sedighmornani et al. (2021) found that shame was positively and significantly associated with insecure, fearful, and anxious attachment styles, whereas individuals with a secure attachment style had lower levels of shame (Gross and Hansen, 2000). A clinical example of the relationship between shame and unresolved and entangled attachment is found in borderline personality disorder (BPD). BPD is characterized by higher levels of baseline emotional intensity for guilt, shame, and fear, and higher levels of shame reactivity (Di Bartolomeo et al., 2022; Estric and Lopez-Castroman, 2022). A “chicken-and-egg” discussion is still ongoing concerning the relationship between attachment and temperament, another possible mediating variable (Hong and Park, 2012). Concerning the Big Five personality traits, e.g., high levels of neuroticism are associated with high levels of experiencing shame, whereas high levels of extraversion are associated with less shame (Erden and Akbag, 2015). Palagini et al. (2018) found that, in turn, insecure attachment in patients with insomnia has been associated with hyperarousal, pre-sleep hyperarousal, and emotion dysregulation. Therefore, clinically, we would expect shame and insomnia of having a mediating role between ACEs and insecure attachment.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

References

- Allen, J. G., Fonagy, P., and Bateman, A. W. (2006). *Handbook of mentalization-based treatment*. Chichester: Wiley. doi: 10.1002/9780470712986
- Bader, K., Schäfer, V., Schenkel, M., Nissen, L., and Schwander, J. (2007). Adverse childhood experiences associated with sleep in primary insomnia. *J. Sleep Res.* 16, 285–296. doi: 10.1111/j.1365-2869.2007.00608.x

Ethics statement

The studies involving human participants were reviewed and approved by the Universiteit van Amsterdam. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FS, EV, and RW contributed equally to this research. JU participated as a graduate student. NN as a clinical expert. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the ZONMW Neuropsychanalysis Fund Grant 16.561.0001 (to FS and EV), and VICI Grant 453.07.001 (to EV), all part of the Netherlands Organization of Scientific Research (NWO); by the Grant 253/2012 of the Bial Foundation (to EV); by the European Research Council Grants ERC-ADG-2014–671084 INSOMNIA and ERC-2016-PoC-737634 INSOMNIA BEAT IT (to EV), and the Australian National Health and Medical Research Council Investigator Grants GNT1196636 (to RW).

Acknowledgments

We thank Dr. A. Lam for providing helpful comments on the description of the statistical methodology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Baglioni, C., Battagliese, G., Feige, B., Spiegelhalter, K., Nissen, C., Voderholzer, U., et al. (2011). Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J. Affect. Disord.* 135, 10–19. doi: 10.1016/j.jad.2011.01.011
- Bastien, C. H., Vallières, A., and Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* 2, 297–307. doi: 10.1016/s1389-9457(00)00065-4
- Bernstein, D. P., Ahluvalia, T., Pogge, D., and Handelsman, L. (1997). Validity of the childhood trauma questionnaire in an adolescent psychiatric population. *J. Am. Acad. Child Adolesc. Psychiatry* 36, 340–348. doi: 10.1097/00004583-199703000-00012
- Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., et al. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse Negl.* 27, 169–190. doi: 10.1016/s0145-2134(02)00541-0
- Blom, K., Jernelöv, S., Rück, C., Lindefors, N., and Kalso, V. (2017). Three-year follow-up comparing cognitive behavioral therapy for depression to cognitive behavioral therapy for insomnia, for patients with both diagnoses. *Sleep* 40:zsx108. doi: 10.1093/sleep/zsx108
- Bruno, A., Rizzo, A., Muscatello, M., Celebre, L., Silvestri, M., Zoccali, R., et al. (2020). Hyperarousal scale: Italian cultural validation, age and gender differences in a nonclinical population. *Int. J. Environ. Res. Public Health* 17:1176. doi: 10.3390/ijerph17041176
- Campbell, J. S. (2016). Development of the military compass of shame scale. *Milit. Behav. Health* 4, 159–167. doi: 10.1080/21635781.2015.1133346
- Christensen, H., Batterham, P. J., Gosling, J. A., Ritterband, L. M., Griffiths, K. M., Thorndike, F. P., et al. (2016). Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. *Lancet Psychiatry* 3, 333–341. doi: 10.1016/S2215-0366(15)00536-2
- DeBellis, M. D. (2010). “The neurobiology of child neglect,” in *The impact of early life trauma on health and disease*, eds R. A. Lanius, E. Vermetten, and C. Pain (Cambridge: Cambridge University Press), 123–133.
- DeYoung, P. (2015). *Understanding and treating chronic shame. A relational/neurobiological approach*. London: Routledge. doi: 10.4324/9781315734415
- Di Bartolomeo, A. A., Varma, S., Fulham, L., and Fitzpatrick, S. (2022). The moderating role of interpersonal problems on baseline emotional intensity and emotional reactivity in individuals with borderline personality disorder and healthy controls. *J. Exp. Psychopathol.* 13:20438087221142481. doi: 10.1177/20438087221142481
- Dorahy, M. J., Corry, M., Black, R., Matheson, L., Coles, H., Curran, D., et al. (2017). Shame, dissociation, and complex PTSD symptoms in traumatized psychiatric and control groups: direct and indirect associations with relationship distress. *J. Clin. Psychol.* 73, 439–448. doi: 10.1002/jclp.22339
- Dorahy, M. J., Middleton, W., Seager, L., Williams, M., and Chambers, R. (2016). Child abuse and neglect in complex dissociative disorder, abuse-related chronic PTSD, and mixed psychiatric samples. *J. Trauma Dissociation* 17, 223–236. doi: 10.1080/15299732.2015.1077916
- Draijer, N. (1990). *Seksuele traumatisering in de jeugd. Gevolgen op lange termijn van seksueel misbruik van meisjes door verwanten. (Sexual traumatization in childhood. Consequences and long term effects)*. Ph D Thesis. Amsterdam: Universiteit van Amsterdam.
- Draijer, N., and Langeland, W. (1999). Childhood trauma and perceived parental dysfunction in the etiology of dissociative symptoms in psychiatric patients. *Am. J. Psychiatry* 156, 379–385. doi: 10.1176/ajp.156.3.379
- Elison, J., Lennon, R., and Pulos, S. (2006). Investigating the compass of shame: the development of the compass of shame scale. *Soc. Behav. Pers. Int. J.* 34, 221–238. doi: 10.2224/sbp.2006.34.3.221
- Estric, C. R., and Lopez-Castroman, J. (2022). Adverse childhood experiences and neurocognition in borderline personality disorder: a call-to-action perspective review. *Harvard Rev. Psychiatry* 30, 248–260. doi: 10.1097/HRP.0000000000000344
- Erden, S., and Akbag, M. (2015). How do personality traits effect shame and Guilt? An evaluation of the Turkish culture. *Euras. J. Educ. Res.* 58, 113–132. doi: 10.14689/ejer.2015.58.4
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences (ACES) Study. *Am. J. Prev. Med.* 14, 245–258. doi: 10.1016/s0749-3797(98)00017-8
- Fonagy, P., Gergely, G., Jurist, E. L., and Target, M. (2002). *Affectregulation, mentalization and the development of the self*. New York, NY: Other Press.
- Germain, A., Buysse, D. J., and Nofzinger, E. (2008). Sleep-specific mechanisms underlying posttraumatic stress disorder: Integrative review and neurobiological hypotheses. *Sleep Med. Rev.* 12, 185–195. doi: 10.1016/j.smrv.2007.09.003
- Griffiths, B. B., and Hunter, R. G. (2014). Neuroepigenetics of stress. *Neuroscience* 275, 420–435. doi: 10.1016/j.neuroscience.2014.06.041
- Gross, C. A., and Hansen, N. E. (2000). Clarifying the experience of shame: the role of attachment style, gender, and investment in relatedness. *Pers. Individ. Dif.* 28, 897–907. doi: 10.1016/S0191-8869(99)00148-8
- Hertenstein, E., Feige, B., Gmeiner, T., Kienzler, C., Spiegelhalter, K., Johann, A., et al. (2019). Insomnia as a predictor of mental disorders: a systematic review and meta-analysis. *Sleep Med Rev.* 43, 96–105. doi: 10.1016/j.smrv.2018.10.006
- Hong, Y. R., and Park, J. S. (2012). Impact of attachment, temperament and parenting on human development. *Korean J. Pediatr.* 55, 449–454. doi: 10.3345/kjp.2012.55.12.449
- Insana, S. P., Kolk, D. J., and Germain, A. (2012). Early-life trauma is associated with rapid eye movement sleep fragmentation among military veterans. *Biol. Psychol.* 89, 570–579. doi: 10.1016/j.biopsycho.2012.01.001
- Jansen, P. R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A. R., et al. (2019). Genome-wide analysis of insomnia (N=1,331,010) identifies novel loci and functional pathways. *Nat. Genet.* 51, 394–403. doi: 10.1038/s41588-018-0333-3
- Kajepeta, S., Gelaye, B., Jackson, C. L., and Williams, M. A. (2015). Adverse childhood experiences are associated with adult sleep disorders: a systematic review. *Sleep Med.* 16, 320–330. doi: 10.1016/j.sleep.2014.12.013
- Kolling, N., Wittmann, M. K., Behrens, T. E., Boorman, E. D., Mars, R. B., and Rushworth, M. F. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* 19, 1280–1285. doi: 10.1038/nn.4382
- Lang, J., McKie, J., Smith, H., McLaughlin, A., Gillberg, C., Shiels, P., et al. (2020). Adverse childhood experiences, epigenetics and telomere length variation in childhood and beyond: a systematic review of the literature. *Eur. Child Adolesc. Psychiatry* 29, 1329–1338. doi: 10.1007/s00787-019-01329-1
- Lanius, R. A., Terpou, B. A., and McKinnon, M. C. (2020). The sense of self in the aftermath of trauma: lessons from the default mode network in posttraumatic stress disorder. *Eur. J. Psychotraumatol.* 11:1807703. doi: 10.1080/2008198.2020.1807703
- Lewis, H. B. (1971). Shame and guilt in neurosis. *Psychoanal. Rev.* 58, 419–438.
- Lichtenberg, J. D., Lachmann, F. M., and Fosshage, J. L. (2011). *Psychoanalysis and motivational systems*. New York, NY: Routledge. doi: 10.4324/9780203844748
- Luyten, P., Envelope, J., and Fontaine, R. J. (2002). Does the Test of Self-Conscious Affect (TOSCA) measure maladaptive aspects of guilt and adaptive aspects of shame? An empirical investigation. *Person. Individ. Differ.* 33, 1373–1387. doi: 10.1016/S0191-8869(02)00197-6
- Manber, R., Edinger, J. D., Gress, J. L., San Pedro-Salcedo, M. G. S., Kuo, T. F., and Kalista, T. (2008). Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disorder and insomnia. *Sleep.* 31, 489–495. doi: 10.1093/sleep/31.4.489
- Morrison, A. (1989). *Shame, the underside of narcissism*. Hillsdale: Analytic Press.
- Muscatello, M. R. A., Rizzo, A., Celebre, L., Mento, C., Pandolfo, G., Cedro, C., et al. (2020). The wounds of childhood: early trauma subtypes, salience and hyperarousal in a sample of adult psychiatric patients. *Int. J. Soc. Psychiatry* 66, 3–9. doi: 10.1177/0020764019872227
- Nicolai, N. (2020). *In levende lijve: het lichaam in de psychotherapie (Dutch ed.)*, 1st Edn. Houten: Bohn Stafleu van Loghum.
- Palagini, L., Petri, E., Novi, M., Caruso, D., Moretto, U., and Riemann, D. (2018). Adult insecure attachment plays a role in hyperarousal and emotion dysregulation in Insomnia Disorder. *Psychiatry Res.* 262, 162–167. doi: 10.1016/j.psychres.2018.01.017
- Pavlova, M., Berg, O., Gleason, R., Walker, F., Roberts, S., and Regestein, Q. (2001). Self-reported hyperarousal traits among insomnia patients. *J. Psychosom. Res.* 51, 435–441. doi: 10.1016/s0022-3999(01)00189-1
- Pesonen, A.-K., Gradsar, M., Kuula, L., Short, M., Merikanto, M., Tark, R., et al. (2019). REM sleep fragmentation associated with depressive symptoms and genetic risk for depression in a community-based sample of adolescents. *J. Affect. Disord.* 245, 757–763. doi: 10.1016/j.jad.2018.11.077
- Porges, S. W. (2011). *The polyvagal theory. Neurophysiological foundations of emotions, attachment, communication and self-regulation*. New York, NY: Norton.
- Pruim, R. H. R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., and Beckmann, C. F. (2015). ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 112, 267–277. doi: 10.1016/j.neuroimage.2015.02.064
- R Core Team. (2021). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Reffi, A. N., Kalmbach, D. A., Cheng, P., Jovanovic, T., Norrholm, S. D., Sexton, M. B., et al. (2022). Sleep reactivity as a potential pathway from childhood abuse to adult insomnia. *Sleep Med.* 94, 70–75. doi: 10.1016/j.sleep.2022.03.026
- Riemann, D., Spiegelhalter, K., Feige, B., Voderholzer, U., Berger, M., Perlis, M., et al. (2010). The hyperarousal model of insomnia: a review of the concept and its evidence. *Sleep Med Rev.* 14, 19–31. doi: 10.1016/j.smrv.2009.04.002
- Riemann, D., Spiegelhalter, K., Nissen, C., Hirscher, V., Baglioni, C., and Feige, B. (2012). REM sleep instability – A new pathway for insomnia? *Pharmacopsychiatry* 45, 167–176. doi: 10.1055/s-0031-1299721
- Schalkwijk, F. (2012). Schaamte in psychotherapie. *Tijdschrift Psychother.* 38, 368–380. doi: 10.1007/s12485-012-0042-9
- Schalkwijk, F., Stams, G. J., Dekker, J., Peen, J., and Elison, J. (2016). Measuring shame regulation: validation of the compass of shame scale. *Soc. Behav. Pers. Int. J.* 44, 1775–1791. doi: 10.2224/sbp.2016.44.11.1775
- Schiel, J. E., Tamm, S., Holub, F., Petri, R., Dashti, H. S., Domschke, L., et al. (2022). Associations between sleep health and amygdala reactivity to negative facial expressions in the UK Biobank cohort. *Biol. Psychiatry* 92, 693–700. doi: 10.1016/j.biopsych.2022.05.023
- Schore, A. N. (2003). *Affect regulation and repair of the self*. New York, NY: Norton.

- Sedighimornani, N., Rimes, K., and Verplanken, B. (2021). Factors contributing to the experience of shame and shame management: adverse childhood experiences, peer acceptance, and attachment styles. *J. Soc. Psychol.* 161, 129–145. doi: 10.1080/00224545.2020.1778616
- Seo, J., Moore, K. N., Gazecki, S., Bottary, R. M., Milad, M. R., Song, H., et al. (2018). Delayed fear extinction in individuals with insomnia disorder. *Sleep* 41:zsy095. doi: 10.1093/sleep/zsy095
- Spinhoven, P., Penninx, B. W., Hickendorff, M., Van Hemert, A. M., Bernstein, D. P., and Elzinga, B. M. (2014). Childhood trauma questionnaire: factor structure, measurement invariance, and validity across emotional disorders. *Psychol. Assess.* 26, 717–729. doi: 10.1037/pas0000002
- Steine, I. M., Skogen, J. C., Krystal, J. H., Winje, D., Milde, A. M., Grønli, J., et al. (2019). Insomnia symptom trajectories among adult survivors of childhood sexual abuse: a longitudinal study. *Child Abuse Negl.* 93, 263–276. doi: 10.1016/j.chiabu.2019.05.009
- Strathearn, L., Giannotti, M., Mills, R., Kisely, S., Najman, J., and Abajobir, A. (2020). Long-term cognitive, psychological, and health outcomes associated with child abuse and neglect. *Pediatrics* 146:e20200438. doi: 10.1542/peds.2020-0438
- Thombs, B. D., Bernstein, D. P., Lobbstaal, J., and Arntz, A. (2009). A validation study of the dutch childhood trauma questionnaire-short form: factor structure, reliability, and known-groups validity. *Child Abuse Negl.* 33, 518–523. doi: 10.1016/j.chiabu.2009.03.001
- Tracy, J. L., Robins, R. W., and Tangney, J. P. (2007). *The self-conscious emotions: theory and research*. New York, NY: Guilford Press.
- Van der Kolk, B. (2014). *The body keeps the score: mind, brain, and body in the transformation of trauma*. London: Penguin.
- van Heukelum, S., Mars, R. B., Guthrie, M., Buitelaar, J. K., Beckmann, C. F., Tiesinga, P. H. E., et al. (2020). Where is cingulate cortex? A cross-species view. *Trends Neurosci.* 43, 285–299. doi: 10.1016/j.tins.2020.03.007
- Voestermans, D., Eikelenboom, M., Rullmann, J., Wolters-Geerdink, M., Draijer, N., Smit, J. H., et al. (2021). The association between childhood trauma and attachment functioning in patients with personality disorders. *J. Pers. Disord.* 35, 554–572. doi: 10.1521/pedi_2020_34_474
- Vonderlin, R., Kleindienst, N., Alpers, G., Bohus, M., Lyssenko, L., and Schmal, C. (2018). Dissociation in victims of childhood abuse or neglect: a meta-analytic review. *Psychological Med.* 48, 2467–2476. doi: 10.1017/S00332911718000740
- Walker, M. P. (2009). The role of sleep in cognition and emotion. *Ann. N. Y. Acad. Sci.* 1156, 168–197. doi: 10.1111/j.1749-6632.2009.04416.x
- Walker, M. P., and Van der Helm, E. (2009). Overnight therapy? The role of sleep in emotional brain processing. *Psychol. Bull.* 135, 731–748. doi: 10.1037/a0016570
- Wang, Y., Raffeld, M. R., Slopen, N., Hale, L., and Dunn, E. C. (2016). Childhood adversity and insomnia in adolescence. *Sleep Med.* 21, 12–18. doi: 10.1016/j.sleep.2016.01.011
- Wassing, R., and D’Rozario, A. (2022). Getting all emotional over sleep. *Health Biol. Psychiatry* 92, 684–686. doi: 10.1016/j.biopsych.2022.08.003
- Wassing, R., Benjamins, J. S., Dekker, K., Moens, S., Spiegelhalter, K., Feige, B., et al. (2016). Slow dissolving of emotional distress contributes to hyperarousal. *Proc. Natl Acad. Sci. U.S.A.* 113, 2538–2543. doi: 10.1073/pnas.1522520113
- Wassing, R., Benjamins, J. S., Talamini, L. M., Schalkwijk, F., and Van Someren, E. J. W. (2019a). Overnight worsening of emotional distress indicates maladaptive sleep in insomnia. *Sleep* 42:zsy268. doi: 10.1093/sleep/zsy268
- Wassing, R., Schalkwijk, F., Lakbila-Kamal, O., Ramautar, J. R., Stoffers, D., Mutsaerts, H. J. M. M., et al. (2019b). Haunted by the past: old emotions remain salient in insomnia disorder. *Brain*. 142, 1783–1796. doi: 10.1093/brain/awz089
- Whalley, M. G., Kroes, M. C., Huntley, Z., Rugg, M. D., Davis, S. W., and Brewin, C. R. (2013). An fMRI investigation of posttraumatic flashbacks. *Brain Cogn.* 81, 151–159. doi: 10.1016/j.bandc.2012.10.002
- Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* 14, 1370–1386.



OPEN ACCESS

EDITED BY

Oliver Hugh Turnbull,
Bangor University, United Kingdom

REVIEWED BY

Daniela Flores Mosri,
Universidad Intercontinental, Mexico
Sérgio Arthuro Mota-Rolim,
Federal University of Rio Grande do
Norte, Brazil

*CORRESPONDENCE

Michael Koslowski
✉ michael.koslowski@charite.de

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 25 October 2022

ACCEPTED 09 January 2023

PUBLISHED 16 February 2023

CITATION

Koslowski M, de Haas M-P and Fischmann T
(2023) Converging theories on dreaming:
Between Freud, predictive processing, and
psychedelic research.
Front. Hum. Neurosci. 17:1080177.
doi: 10.3389/fnhum.2023.1080177

COPYRIGHT

© 2023 Koslowski, de Haas and Fischmann.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Converging theories on dreaming: Between Freud, predictive processing, and psychedelic research

Michael Koslowski^{1,2*†}, Max-Pelgrom de Haas^{1,2†} and
Tamara Fischmann^{2,3}

¹Department of Psychiatry and Psychotherapy, Charité CCM—Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, ²Clinical Psychology and Psychoanalysis, International Psychoanalytic University Berlin, Berlin, Germany, ³Focus III: Psychotherapy and Psychoanalytic Conceptual Research, Sigmund-Freud-Institut, Frankfurt am Main, Germany

Dreams are still an enigma of human cognition, studied extensively in psychoanalysis and neuroscience. According to the Freudian dream theory and Solms' modifications of the unconscious derived from it, the fundamental task of meeting our emotional needs is guided by the principle of homeostasis. Our innate value system generates conscious feelings of pleasure and displeasure, resulting in the behavior of approaching or withdrawing from the world of objects. Based on these experiences, a hierarchical generative model of predictions (*priors*) about the world is constantly created and modified, with the aim to optimize the meeting of our needs by reducing prediction error, as described in the *predictive processing* model of cognition. Growing evidence from neuroimaging supports this theory. The same hierarchical functioning of the brain is in place during sleep and dreaming, with some important modifications like a lack of sensual and motor perception and action. Another characteristic of dreaming is the predominance of *primary process thinking*, an associative, non-rational cognitive style, which can be found in similar altered states of consciousness like the effect of psychedelics. Mental events that do not successfully fulfill an emotional need will cause a prediction error, leading to conscious attention and adaptation of the priors that incorrectly predicted the event. However, this is not the case for *repressed priors* (RPs), which are defined by the inability to become reconsolidated or removed, despite ongoing error signal production. We hypothesize that Solms' RPs correspond with the *conflictual complexes*, as described by Moser in his dream formation theory. Thus, in dreams and dream-like states, these unconscious RPs might become accessible in symbolic and non-declarative forms that the subject is able to *feel* and make sense of. Finally, we present the similarities between dreaming and the psychedelic state. Insights from psychedelic research could be used to inform dream research and related therapeutic interventions, and vice versa. We propose further empirical research questions and methods and finally present our ongoing trial "Biological Functions of Dreaming" to test the hypothesis that dreaming predicts intact sleep architecture and memory consolidation, *via* a lesion model with stroke patients who lost the ability to dream.

KEYWORDS

dreaming, predictive processing, prior, dream coding, psychedelics, psychoanalytical, consciousness

1. Introduction: Freud, homeostasis, feelings, and predictions about the world

When Freud stated in 1894 that “quotas of affect spread over the memory-traces of ideas somewhat as an electric charge is spread over the surface of a body” (Freud, 1894), he laid the ground for what is now common knowledge in affective neuroscience that arousal processes arise in the brainstem and are felt as affects which are distinct from memory-traces of ideas (Solms, 2013). The latter are representational processes that involve forebrain and cortical processes. According to Freud, the distinction between representational processes and *quotas of affects* that are activated by these arousals lays the ground for how the brain works, namely, by feelings. Consciousness registers the state of the subject by feelings and not that of the object world, namely, perception. Freud went on to consider: “If now we apply ourselves to considering mental life from a biological point of view, an instinct appears to us as a concept on the frontier between the mental and the somatic, as the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demand made upon the mind for work in consequence of its connection with the body” (Freud, 1915). This points to how we understand brain processes that give rise to pleasant and unpleasant affects in relation to the mechanism of homeostasis.

Freud thought that we must turn to biology because instincts [Triebe] are fundamentally biological processes. According to his theories, the fundamental developmental task is to learn how to meet our needs in the world and how to manage our emotional needs. The homeostasis principle can guide us in this regard by introducing a value system giving direction if we approach or withdraw from the world of objects, to maintain a state of homeostasis where our needs are met (Solms, 2019). This value system is pleasure and *unpleasure* (emotional pain) as affective signals that are felt when our needs are met or not. When Freud realized from his clinical observation that humans are not only searching for pleasure but are also looking for something “deeper,” beyond pleasure, he called it “the Nirvana principle” (Freud, 1920): no needs, no demands upon the mind, which he distinguished from the pleasure principle. But now, we know that they are not two separate principles; the pleasure principle is *in the service of* the Nirvana principle. They are one and the same principle called the homeostasis principle. The state of Nirvana conceptualized from the perspective of the principle of homeostasis is the state of no need, which can biologically be seen as the “ideal life situation” (Carhart-Harris and Friston, 2010; Solms, 2018). Whether our prediction of how to meet our needs is accurate or not, this is the work the mind has to perform in the service of homeostasis to stay within our viable bounds. *Feelings* can be described as the mental and affective representation of how we maintain homeostasis. We are born with certain predictions of how to meet our needs, and those innate predictions are called reflexes and instincts. Reflexes and instincts yet seem to lack the complexity to meet all human needs. They are automatic, stereotyped responses, while humans need to behave and react in different, adaptive ways depending on the situation in question. We are required to learn from experience how to satisfy our needs. This means we must supplement our innate predictions as they are too crude. We need more context-related methods to satisfy our drives. That is the whole

task of ego-development, of learning from experience, and of building an internalized representation of how the world works because it is in the world where we have to satisfy our drives. The psychoanalytic *Ego* is all about predictions. Its task is to learn from the experience of the past to predict how to go about meeting our needs in the future (Carhart-Harris and Friston, 2010; Solms, 2018). To the extent that we learn predictions that are not adequate and do not work, to that extent we suffer from feelings.

2. Predictive processing and the dreaming brain

Dreaming always had a strong significance in Freud’s core concepts of human consciousness, the ego, and the unconsciousness; he considered dreams the “royal road to the unconscious” (Freud, 1900). Many of Freud’s initial theories have been successfully updated and aligned with modern psychological and neurobiological frameworks (Solms, 2018). One currently influential model within computational neuroscience and epistemology is the predictive processing (PP) account. We will demonstrate in the following analysis and comparison of different theories and aspects from different perspectives that the PP model can reasonably be applied to a psychoanalytical understanding of dreaming and that promising insights and new hypotheses can be derived from this cross-theoretical approach.

The PP or predictive coding theory of brain functioning has its roots in Bayesian probability statistics. Bayesian statistics describe the uncertainty of events in terms of their mathematical probability, meaning the degree of belief in an event based on up to that point acquired knowledge (Spiegelhalter and Rice, 2009). The PP model offers a unifying framework for imagination, perception, and the brain’s organization as a sense-making-organ, which tries to find causation within the world that surrounds it (Clark, 2012; Hohwy, 2013). It postulates that the brain can be understood as a hierarchically structured inference system, operating under the premise of minimizing long-term prediction error. Within these multilayered structures, each layer generates expectations, the so-called priors, of what sensory information it will receive by the next layer below. Depending on the layer’s hierarchical position, these priors vary in their degree of abstraction and temporal reference. Bottom layer predictions deal with detail-rich, fast-scale perceptual elements, getting progressively oriented toward long-term meta-features at the top. Predictions are then matched with the actual bottom layer input, resulting in a quantifiable prediction error depending on the discrepancy between input and expectations. Over time, the system will enhance the accuracy of its priors to make an adequate assumption about internal and external conditions. Depending on the environment and the task to be accomplished, the weighting of prediction error signals can be dynamically adjusted. For example, in environments in which sensory signals are weak or highly fault-prone (“noisy” conditions, like visual perception in darkness), the dependency on higher-level priors is increased, while lower-level priors are attenuated. PP’s synthesis of active “world-creation” and passive “world-sensing” enables an integrated understanding of mental operations previously conceptualized rather separately. As Clark expressed: “Perceivers like us ... are inevitable potential dreamers and imaginers too” (Clark, 2012).

Since the PP model gained momentum in cognitive neuroscience, there have been attempts to apply the PP account not only to waking consciousness but also to dream. Clark, Hobson, and Friston argue that the same hierarchical functioning is in place during sleep, with the exception of a lack of sensitive and motor perception and action (Crick and Mitchison, 1983; Hobson and Friston, 2012; Hobson, 2014; Clark, 2016). For a more detailed review of the application of the PP account on sleep and dreaming, refer to the study by Bucci and Grasso (2017). According to the authors, REM sleep forces the brain to exclusively rely on its middle-to-high-level priors to deal with random neuronal activity since low-level priors are evaluated as unreliable. This explains the dreamscape's lack of "the fine-grained perceptual details and depth" compared to waking life (Bucci and Grasso, 2017). While this explanation relies predominantly on the neural signature of REM sleep, there is growing evidence that dreaming also happens during NREM sleep and that there might be continuity between different states of consciousness (e.g., waking, daydreaming, and dreaming) involving complex cortical activation patterns (i.e., the default mode network [DMN]), which does not contest the basic PP hypothesis for dreaming (Domhoff, 2011; Fox et al., 2013).

3. From Freud to Friston: Primary process and free energy principle

The distinction between *primary* and *secondary* processes is a classic idea of Freudian psychoanalysis (Freud, 1900). It accounts for the fundamentally different modes of cognition between ordinary, adult waking consciousness, and non-ordinary states such as dreaming, psychosis, or infantility. While the former is characterized by an ordered, rational, and coherent style of mental operations, impulse control, and reality testing, the latter is described as the persistence of more primitive and chaotic forms of thinking, essentially being a regressive, developmentally outgrown type of *protoconsciousness*. According to Freud, the primary process describes the cognitive style of the *id/unconscious*, and the secondary process is the one of the *ego* or, in less psychoanalytical terms, the cognitive functioning of a healthy, rational human mind in its waking state (Freud, 1940). In this framework, dreams are generated in the pre-verbal, unconscious space of the primary process. For a more thorough, critical discussion of the concept of the primary process and its relation to primordial consciousness and repression, refer to the study by Robbins (2018).

In the states of sleep and dreaming, the same dopaminergic SEEKING system (Panksepp and Wilson, 2016; capital letters by the authors) is active as in waking life. However, in sleep, the sensory input is reduced to a minimum. This makes the brain free to minimize complexity in REM dreaming and hence assumedly resolves otherwise disruptive strong emotional reactions, which could interrupt continuous sleep. In psychoanalytic dream theory, it is assumed that impingements from inside (emotional needs) or outside (life events and day residues) are major causes and sources for dreaming (Freud, 1900).

This aspect has found access to Friston's concept of free energy (FE) (Friston, 2012). Friston hypothesizes that the brain operates to minimize FE caused by sensory impingements of unpredicted stimuli. Similar to Freud who assumed major needs or biological imperatives to reflect such impingements, Friston proposes that

these sensory impingements reflect compliance with biological imperatives, creating a "demand for work" (Freud, 1915) to produce "specific actions" or, to put it in Friston's words, "an imperative to minimize prediction error ... through action" (Friston, 2012). For both Freud and Friston, when such sensory impingements are felt, they put a demand on our brain to embody a representation of these sensory impingements including representations of the bodily ego (Freud, 1923) or as Friston would term it the "agent's body." These bodily representations are initially met by an innate generation of a prior virtual version of reality, i.e., a constructed and simplified version of reality, which will subsequently be modified by experience.

We might say from here that Freud's primary process and Friston's "virtual reality generator" can be seen as innate producers of imaginary prior beliefs and predictions of the actual experience. According to Freud, the primary process is "in the apparatus first," which could be termed in the PP model as higher-order priors or beliefs that are usually not accessible to consciousness, hence unconsciousness (Freud, 1900).

According to other influential dream theories (threat simulation theory and social simulation theory), dreaming can be described as a perceptual synthesis by testing real-life experiences in a virtual setting, co-creating, and updating a generalizable generative model of the world, in order to simulate threatening or important social situations (Revonsuo, 2000; Ruby, 2011; Tuominen et al., 2019; Scarpelli et al., 2022). The generative model is thereafter tested in conscious waking life for its feasibility and precision of the prior predictions, producing prediction errors and, thus, "surprise," change in behavior, and further model updating, when real-life perceptions differ from the predictions.

In conclusion, the PP account is well compatible with both traditional psychoanalytical dream theory and more recent cognitive dream theories.

4. The repressed prior

Mark Solms' core conception called "predictive processing and the feeling brain" is a Neo-Freudian, neurobiologically informed model based on the principle of the homeostasis of feelings, which is governed by the pleasure principle (Solms, 2018, 2021). For our focus on dream research, the unconscious is of particular interest. In Solms' conception, the unconscious consists of different functional subsystems, which can be mapped onto distinct brain networks:

1. The "system unconscious," where "the repressed" is derived from cognitive (representational) processes, acquired by non-declarative learning. Its functions are performed by subcortical brain structures (basal ganglia, cerebellum).
 - Cognitive unconscious: *legitimately* (maturely) automatized predictions; the normal case, because "predictions work well".
 - Dynamic unconscious: *illegitimately* (prematurely) automatized predictions, the repressed; the pathological case, because the need could not be satisfied (e.g., repressions due to the Oedipus complex) → *the repressed prior* (RP).
2. The "preconscious," consolidated declarative memory content, is performed by the cortex.

Following this model, mental events that do not successfully fulfill an emotional need cause a prediction error, leading to conscious attention, problem-solving, consecutive reconsolidation and adaptation of the priors that predicted the event. However, this is not the case for *repressed priors*, which are defined by the inability to become reconsolidated or removed, despite ongoing error signal production (Solms, 2018). The RP is described as a prematurely formed, maladaptive automatization of the infantile prediction of an apparently insolvable problem (conflict). There is no explicit mental representation of an RP that could be experienced or verbalized, and its impact remains in the unconscious, affective layers of consciousness. This automatization might follow the economic rationale that a better solution cannot be found at the current moment and a conscious re-engagement with the problem would occupy too much of the capacities of working memory, thus preventing it from dealing with issues it can actually solve (Solms, 2018).

Interestingly, the concept of the RP shares some core aspects with the *unsolved conflictual complexes*, as described by Moser (Moser and von Zeppelin, 1996; Moser and Hortig, 2019), which is presented in the following section. By doing so, we intend to bridge a gap between the PP account of dreaming and the psychoanalytically inspired dream generation model of Moser et al.

5. The psychodynamic dream generation model

Moser's dream generation model (Moser and von Zeppelin, 1996; Moser and Hortig, 2019) is based on psychodynamic dream theory, developmental and cognitive psychology, as well as experimental dream research. Moser et al. consider the sleep dream as a simulated *micro-world* controlled by affectivity, which generates images of entities involved in it and their relations to each other. A dream is triggered by current concerns of events that happen during the day (day-residues), which are capable of reactivating unresolved conflicts and problems due to structural similarities. The dream having the function to resolve those conflicts can do so more readily in contrast to the waking state, as the dream state has no capacity restrictions of the memory system. Consolidation processes can also take place during sleep in the so-called *off-line mode*. This is how new information is integrated into long-term memory while sleeping. The range of affect modulation is significantly larger in the micro-worlds of dreaming and stress is absorbed readily both *via* imagination and cognition. The dream is not involved in regulating concrete-real object relationships but rather works with memories, acquired solutions, and defense strategies, which are regrouped as *prototypical affective microprocesses* (PAMs).

A dream, which is usually pictorial, consists of at least one situation produced by a *dream-organizer*. According to Moser, a dream-organization may be considered a bundle of affective-cognitive procedures (i.e., PAMs), generating the micro-world dream and controlling its course of action. Within this system, the so-called *dream-complex* serves as a template for facilitating dream organization. Thus, it may be assumed that a dream-complex originates from one or more complexes stored in long-term memory, rooted in conflictual or traumatizing experiences, which found their condensates in introjects. These conflictual or traumatic dream complexes are easily triggered by stimuli from the outside world, which are structurally similar to stored situations of these complexes.

Searching for a solution for this complex is governed by the need for security and wish for involvement, i.e., the *security-principle* and the *involvement-principle* which govern dream-organization.

Within these complexes, wishes are the links between PAMs of self and objects, which are accompanied by cognitive aspects such as convictions and hope for wish-fulfillment. *Conflictual complexes* are areas of PAMs with a repetitive character, thus creating areas of unbound affective information. Affects within such an area are inter-connected but blocked from memory and, thus, not accessible for our conscious cognition. They are the integrated affects, which due to their lack of representational contextualization can hardly be modulated or resolved, and that makes the patient suffer from conflictual complexes. To solve these conflictual or traumatic complexes, it is necessary to retrieve the affective information and re-integrate them into a relational reality to make the complex "come alive." This is being attempted in dreams whose function is to search for a solution to the complex. The search for a solution within a dream is governed by the abovementioned need for security and a wish for involvement.

In order to collect empirical data about the dream generation described in Moser's theoretical framework, a standardized method to codify the manifest dream was developed, the *Zurich Dream Process Coding System* (ZDPCS). It enables to scientifically measure alterations in dream contents, which has been shown in several empirical studies (Fischmann et al., 2013, 2021; Fischmann and Leuzinger-Bohleber, 2017; Wittmann et al., 2017).

We will later propose that the use of instruments like the ZDPCS would be an interesting tool to investigate also other, dream-like states of waking consciousness, e.g., the psychedelic experience, which is the state of the brain and consciousness during the influence of a serotonergic (5HT-2a receptor agonist) substance. Those states can be used as a promising experimental condition to better understand the process and structure of dream generation, from both a phenomenological-psychological and a neurobiological perspective. On the neurobiological level, PP could be the best available framework to link these phenomenological data with cognitive neuroscience methods and paradigms.

6. Repressed priors and conflictual complexes: Bridging the two models

We discussed earlier that RPs, according to Solms, are automated response patterns of the non-declarative long-term memory system. They constantly produce prediction error signals, which can be "felt" as displeasure, but they cannot be "thought" as they lack properties of cognitive representation. In psychodynamic terms, they are repressed to the unconscious. The RPs are described as not having a representable form and, thus, could not become reconsolidated whatsoever. We hypothesize that Solms' RP corresponds to the conflictual complex (CC) in Moser's theory of dream generation, i.e., the RP/CC is considered one single entity, described from two different angles. See Table 1 for a comparison of the core features of RP and CC. We agree though that an RP/CC could possibly never be expressed *directly* in a cognitive representation—one needs to dream, or enter a dream-like state, to make the unbound affect (according to Moser)/unfulfilled emotional need behind the error accumulation (according to Solms) accessible to further processing in a represented manner.

TABLE 1 Repressed prior and conflictual complex—comparison of core features.

	Repressed prior (Solms)	Conflictual complex (Moser)
Features	Unconscious	Blocked from consciousness
	No cognitive representation	Association to current concern elicits dream
	Automatized response	Unbound affective information
Origin	Repressed due to unsolvable conflict	Repressed due to unresolved conflict
	Acquired in early Development	
Dynamics	A priori non-representable (only the affective part)	Associated to current concern making it accessible
		Access possible (e.g., by dream work in psychotherapy)

Thus, we propose that in line with Moser's dream theory and Solms' neuropsychodynamic framework, the underlying emotional need, which these priors attempt but fail to satisfy, can become contextualized and experienceable *within dreams and similar states of experience and neurobiological functioning*. Dreaming might be a state of cognition in which these unresolved emotional needs can be approached with alternative solutions for wish-fulfillment and conscious access. In the PP model, this would mean that the affective consequence of the malfunctioning RP shapes the conscious experience of the dream, e.g., in the form of an affect-laden symbol, with semantic associations between dream content and self-narratives. The following processing of the transformed content during the waking state, e.g., generating personal meaning of a bizarre dream like one does in psychoanalytical dream interpretation, might then result in new insights about maladaptive (pathological) patterns of prediction of inner and outer world events, and potentially reduce cognitive, affective, and behavioral rigidity. This dynamic process might facilitate new predictions and shape behavioral reactions that are more capable of meeting unmet emotional needs. The dream state might thus continue and potentiate processes that might have been initiated previously during waking consciousness. To account for the proposed function of dreams, they need to be understood as presenting a categorically different "mode of operating" or "style of cognition" than that exhibited by the brain during ordinary waking consciousness—primary process thinking, as opposed to secondary process thinking, as we stated earlier.

We hope that by understanding dreaming as a primary process, we can build a hypothesis to account for two things. First, dreams can incorporate the conflicted complex in ways waking consciousness cannot. Second, how the dream state might facilitate alternative ways of solving these complexes (unfulfilled emotional needs) to settle in.

We hypothesize that with the weakening of ego-control and its secondary defenses, the prediction error of the repressed prior can become felt as the unmet emotional demand it fails to fulfill. On a neurobiological level, one possible explanation might be that as the default-mode network's (DMN)¹ control over the medial temporal

lobe (MTL)² decreases within primary consciousness (DMN-MTL decoupling, Carhart-Harris et al., 2014), affective impulses might then activate associatively connected contents of the declarative memory systems in a bottom-up manner and become integrated *via* the association cortices around the TPO-region (Solms, 2000) into the embodied, simulated micro-world of the dream state.

We hypothesize that with the strong restriction of external sensory input and the decreased precision weighting of high-level priors, the generative freedom for action and scenarios beyond the rules of everyday reality and entrenched reaction patterns substantially increases. A significant function of high-level priors is the stratification of inferior levels within the inferential hierarchy, channeling or streamlining predictive roads³, which are the most in line with the core assumptions of the system. The result is a constrained amount of prioritized, pre-structured predictive paths (the metaphor of predictive highways might be suitable) in accordance with Occam's principle, which enables efficient functioning (control and agency) during waking activity. These earlier prioritizations may either be legitimate, meaning that they have proven to work well in fulfilling the need they serve and have been, therefore, correctly automatized (Solms' cognitive unconscious) or illegitimate in the context of what has been called the dynamic or repressed unconscious (Solms, 2018). Again, in this terminology, repression means that this prediction remains protected from reconciliation (*via* precision weighting) despite its production of error signals.⁴

With the reduced weighting of higher-level priors, their restrictive control on lower levels diminishes and gives rise to less pre-emphasized and combinatory unrestrained mid-level priors. Sticking to our metaphor, our highways have turned into normal roads and resemble just one option besides others which are now more

2 The MTL including structures such as the hippocampus, amygdala, and para-hippocampal regions plays a central role in the functioning of the declarative memory system (Cutsuridis and Yoshida, 2017). Damage of these structures does not cause secessions of the ability to dream all together (Solms, 2000), but does impair the expressiveness of dream reports. For example, bilateral hippocampal lesions have been reported to reduce the episodic quality and content-richness of dreams (Spanò et al., 2020), and dreams of patients with bilateral damage of the amygdala appeared to be significantly shorter and less complex (Blake et al., 2019).

3 We chose the term "predictive roads" to illustrate that the deeply consolidated priors at the core of the predictive hierarchy (non-declarative LTM priors in form of "action tendencies") channel or orchestrate the consecutively following predictions on inferior levels such as motor-programs, mental images, declarative contents, and so on. This is how our meta-priors choose or direct the downstream routes our predictive cascades may take.

4 Importantly, Solms proposed the conception of repression with regard to unsolvable problems argues along a similar mechanism as Carhart-Harris' proposition for adverse life events (Carhart-Harris, 2019): 'The mnemonic Tightened Beliefs in Response to uncertainty (TIBER) is offered to capture this phenomenon. Here "trauma" is interchangeable with "uncertainty"—if used in an extended sense to mean any significant acute and/or repeated adversity that is paralleled by significant uncertainty. The basic tenet is that (implicit) beliefs tighten as a defensive response to significant, intolerable stress and uncertainty.' Insofar, one might suspect that Moser's 'conflicted, traumatic complex' also seem to root in a consistent underlying mechanism of aberrant precision-weighting.

1 According to Carhart-Harris and Friston (2010), the DMN represents the neurological correlate of the psychoanalytic ego.

accessible. This in turn, as stated previously, might facilitate newly learned, more mature predictions to be tried out in the “safe space” of the dream-world and continue to settle in.

7. Neural dynamics of the primary process: Data from research with psychedelics

Recently, with the increase in neuroscientific research on psychedelic states, the concept of primary and secondary process thinking has regained academic interest and popularity. A pioneering study by Carhart-Harris et al. aligned these conceptions with the predictive brain theory and the entropic brain hypothesis (Carhart-Harris and Friston, 2010; Carhart-Harris et al., 2014), arguing that the essential quality of the secondary process is to minimize free-energy (entropy, uncertainty) *via* top-down predictions which suppress occurring prediction errors on lower levels of the inferential hierarchy. They provide empirical data that imply that psychedelics disrupt these processes, particularly neural activity in the DMN, and induce a primary state of consciousness that is hypothesized to have specific underlying neurophysiological characteristics. This disruption of the hierarchical predictive architecture and a compromised capacity of top-down control results in less constrained cognitions and more chaotic (higher-entropy) neural dynamics, which on the subjective level goes along with primary process thinking. Summarized in the REBUS formulation (“relaxed beliefs under psychedelics”), the authors argue that in these states, high-level priors become deemphasized *via* reduced *precision-weighting* and allow for a broader range of lower-level activity to occur, explaining the distinct phenomenology of the psychedelic experience (Carhart-Harris and Friston, 2019). Other research groups stressed different aspects of brain activation patterns, e.g., dysfunction of thalamic gating, leading to very similar conclusions for subjective experience and brain functioning in the psychedelic state (Vollenweider and Kometer, 2010; Preller and Vollenweider, 2018).

8. Dreaming and the psychedelic experience

It has been stated repeatedly that the phenomenology of dreaming shares many similarities with the subjective experience after ingestion of a psychedelic substance with strong agonism at the serotonin (5-HT)-2A receptor, such as psilocybin, LSD, mescaline, or DMT/Ayahuasca (Schultes and Hofmann, 1979; Kraehenmann, 2017; Palhano-Fontes et al., 2021). A study comparing dream reports and reports of psychedelic experience found perceptual changes and close relationships as the most prevalent themes in both conditions (Sanz et al., 2018). The phenomenology of the psychedelic experience has been assessed and mapped by the use of several standardized questionnaires and qualitative interviews in relation to the dose of the applied psychedelic substances (Griffiths et al., 2006; Studerus et al., 2011; Millière, 2017; Preller and Vollenweider, 2018; Holze et al., 2021). The available quantitative data in relation to the applied dose of psychedelics have recently been regrouped in the Altered States Database project (Schmidt and Berkemeyer, 2018).

In a study with healthy subjects performing a mental imagery task after the ingestion of a high dose of the psychedelic LSD,

the authors observed a shift of subjective experience toward the abovementioned primary process thinking (Kraehenmann et al., 2017). The use of the term “primary process thinking” here follows the conceptualization of Auld (Auld et al., 1968), who developed a scale for the evaluation of dream reports. This scale sums up scores in nine categories, namely, condensation, unlikely combinations or events, fluid transformations, visual representation, symbolism, contradiction, magical occurrences, inhibited movement, and taboo sexual and aggressive acts. These elements were then related to secondary process thinking, to acquire the *primary index* (PI), as established in studies on guided mental imagery and daydreaming (Stigler and Pokorny, 2001). The authors concluded that both dreaming and the psychedelic state share a distinct mode of cognition characterized by primary process thinking.

Concerning the experiential domain of the sense of self in altered states of consciousness such as sleep, dreaming, meditative states, and the psychedelic experience, there is a growing body of literature bridging core concepts from the philosophy of mind and cognitive science (Letheby and Gerrans, 2017; Millière et al., 2018). In the case of the psychedelic experience as a “lab model” for the study of altered states of consciousness, there is convincing empirical evidence in favor of a stepwise disintegration in the sense of a coherent, enduring, temporally and spatially well-defined self-as-object. This process can be seen as *self-unbinding* (Letheby and Gerrans, 2017), extending the model of cognitive binding (Sui and Humphreys, 2015) to explain the psychedelic-induced loss of ego functions and body boundaries. This model is simultaneously informed by the abovementioned account of PP and the REBUS model, allowing empirical hypothesis testing with neuroimaging methods. This seems a fruitful direction to further investigate changes in self-experience in dream research.

Comparing the phenomenology of dreaming and the experiences induced by high doses of a psychedelic⁵, the following commonalities and differences can be described.

The similarities in phenomenology between dreaming and the psychedelic experience are obvious and manifold, also with more fine-grained distinctions of the different features (Edwards et al., 2013; Kraehenmann, 2017; Kraehenmann et al., 2017; Letheby and Gerrans, 2017; Schmidt and Berkemeyer, 2018). Table 2 shows strong similarities in the domains of perception, affect, and cognition.

Regarding the differences, we are inclined to locate the psychedelic state closer to lucid dreams (LD) than regular (REM-) dreams, as in LD meta-cognition, reality monitoring and memory functions are preserved, conscious choices can be made, and real-time communication is possible *via* eye movements (Dresler et al., 2014; Baird et al., 2019; Mota-Rolim, 2020; Loo and Cheng, 2022). Real-time communication offers new possibilities for empirical research and improves the precision and richness of subjective reports, as research on regular dreams almost always suffers from the indirect, *a posteriori* nature of dream reports.

The specific feature of elementary visual imagery in the psychedelic state (e.g., circles, triangles, colored patterns on object surfaces, halo effects) might be due to the neurobiological feature of the 5HT-2A-receptor agonism of psychedelics, inducing changed sensory drive and temporal dynamics in the visual cortex (V1) which in turn lead to impaired integration of visual perception

⁵ Most of the recent clinical trials use Psilocybin 25 mg or LSD 100–150 µg as a single high dose to induce the full spectrum of a psychedelic experience.

TABLE 2 Phenomenology of dreaming and the psychedelic experience.

	Domains of phenomenology				
	Perception	Affectivity	Cognition	Sense of self	Behavior
Shared phenomenological features: Dreaming and psychedelic experience	Vivid, predominantly visual imaginary perceptive changes; bizarreness; symbol formation	Strong activation of emotional memories and affects (positive and negative valence); retrieval of fear memory: nightmare/“challenging experience”, “bad trip”	Decrease of logical and increase of associative reasoning; shift toward bizarre, symbolic and metaphoric thinking; insightfulness; primary process thinking	Disintegration of narrative and embodied (minimal) self, non-dual awareness, depersonalization/derealization	
Differences in phenomenology					
Dreaming	REM dreams: mostly complex images; influence of external stimuli is marginal (“slamming door”)		No meta-cognition/reality monitoring; memory functions partly preserved; exception: lucid dreams		Lack of motor control; exception: lucid dreams (voluntary control of eye movements)
Psychedelic experience	Complex images and elementary percepts (abstract geometrical forms) mental imagery modified by external perception (e.g., synesthesia)		Metacognition/reality monitoring and memory functions mostly preserved		Motor control mostly preserved

(Michaël et al., 2019). This difference might limit the comparability of the two states.

In dreams, there is an obvious lack of external stimulation, whereas, in the psychedelic state, external stimuli are known to have a substantial influence on the subjective experience. This difference might play a minor role in current empirical research, as in the standardized experimental setting of modern clinical trials with psychedelics, the subjects wear eyeshades, have their eyes closed, and are asked to direct their attention toward the “inner world,” focusing on arising images, feelings, thoughts, and somatic experiences—the “outer world” stays mostly outside (Koslowski et al., 2021). Still, the fact that many trials include listening to emotionally activating music might be a source of bias in the direct comparison of these states.

The lack of motor control in dreams seems to be another important difference, from the perspective of embodied mind regarding motor-proprioceptive feedback loops, while we should bear in mind that brainstem atonia in REM sleep is not as complete as one might think (Windt, 2015).

Taken together, we think that the phenomenological similarities between dreaming and the psychedelic state outweigh by far their differences. Particularly, there are many structural similarities from the perspective of the visual, affective, and cognitive features.

9. Discussion

In the sections 1 to 6, we discussed how the different features of dreaming can be reasonably described in terms of the PP model of cognition and that a current theory of dream generation (Moser) is well compatible with this view. We argued for the commonalities of Solms’ repressed prior and Moser’s conflictual complex, which could help inform both theories with testable hypotheses. In section 7, we introduced the neurocognitive perspective on the psychedelic experience, which, like dreaming, has been explained in terms of the PP model. In section 8, we reviewed the phenomenological similarities and

differences between dreaming and the psychedelic experience and proposed the application of a psychedelic substance as an experimental model to induce a dream-like state in waking consciousness, which would allow us to expand the limits of dream research due to improved perception and communication of the actual experience.

Following the arguments in the sections earlier, we derive the following propositions for future dream research:

1. The earlier mentioned dream coding system ZDPCS could be used in empirical trials to further investigate the described similarities in the formal structure of experience in dreaming, normal waking state, and dream-like waking states, e.g., pharmacologically induced by the application of psychedelic substances.
2. Insights from psychodynamic dream work could be therapeutically relevant for psychedelic-assisted therapy, as there is still an ongoing discussion on which psychotherapeutic model one should apply (Wolff et al., 2020; Koslowski et al., 2021; Yaden et al., 2022). The repressed priors/conflictual complexes of the PP account of dreaming could possibly be “traced,” reconstructed, and reconciled, at least their affective part, using specific psychodynamic interventions (i.e., free association, symbolization, and psychodynamic interpretation of the experienced content).
3. Functional neuroimaging paradigms and other neuroscience experiments from research on psychedelics could be used in different sleep stages (REM/non-REM) and dream conditions (regular dreaming/lucid dreaming) to further elucidate the mechanisms of dream formation. This is particularly interesting for paradigm testing for the different layers of the PP account (e.g., visual system, emotion processing, higher-order beliefs, and sense of self).
4. Lesion models that mimic the loss of the ability to dream could shed light on the proposed biological and evolutionary functions of dreaming, as stated by Freud and Solms. One could examine patients with a distinct brain lesion associated with the ability to dream to further investigate some of the above-described hypotheses.

The fourth proposal in this list, brain lesion models for dream research, is about to be realized in our ongoing observational trial “BFD—Biological Functions of Dreaming” (<https://clinicaltrials.gov/ct2/show/NCT04749992>). This project aims to understand the biological function of dreams, which differs from that of REM sleep. Based on Mark Solms’ neuropsychanalytical theory and neuropsychological findings that REM sleep and dreams are doubly dissociable phenomena (Solms, 2000, 2014), Freud’s central hypothesis that dreams serve to maintain sleep will be investigated further (Freud, 1900). By this, Freud meant that the dream is a response to affect-laden impulses for action with hallucinatory wish-fulfillment so that it does not lead to awakening. Second, it will be examined whether dreams influence the consolidation of affective and non-declarative (motor) memory. Our hypotheses are that patients who have lost the ability to dream during REM sleep have poorer sleep quality and poorer emotional and non-declarative memory consolidation. This will be investigated in two groups of neurological patients with thrombotic infarction in the area of the posterior cerebral artery (PCA) who have lost the ability to dream while retaining REM sleep. The PCA stroke was selected as a lesion model because it has been shown that a lesion in the temporoparietal junction and related structures in the PCA area frequently led to a loss of dreaming (Solms, 2014).

We hope that our considerations might inspire other researchers to take further steps in the interdisciplinary terrain of empirical research on dreaming and altered states of consciousness.

Author contributions

MK and M-PH conceived the central theoretical ideas presented in this article. TF contributed to Mosers theory of dreaming and provided critical feedback. MK, M-PH, and TF had equal contributions and wrote the manuscript.

References

- Auld, F., Goldenberg, G. M., and Weiss, J. V. (1968). Measurement of primary-process thinking in dream reports. *J. Pers. Soc. Psychol.* 8, 418. doi: 10.1037/h0025488
- Baird, B., Mota-Rolim, S. A., and Dresler, M. (2019). The cognitive neuroscience of lucid dreaming. *Neurosci. Biobehav. Rev.* 100, 305–323. doi: 10.1016/j.neubiorev.2019.03.008
- Blake, Y., Terburg, D., Balchin, R., van Honk, J., and Solms, M. (2019). The role of the basolateral amygdala in dreaming. *Cortex* 113, 169–183. doi: 10.1016/j.cortex.2018.12.016
- Bucci, A., and Grasso, M. (2017). *Sleep and Dreaming in the Predictive Processing Framework*.
- Carhart-Harris, R. (2019). How do psychedelics work? *Curr. Opin. Psychiatry* 32, 16–21. doi: 10.1097/YCO.0000000000000467
- Carhart-Harris, R., and Friston, K. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010
- Carhart-Harris, R., and Friston, K. (2019). REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. *Pharmacol. Rev.* 71, 316–344. doi: 10.1124/pr.118.017160
- Carhart-Harris, R., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., et al. (2014). The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 20. doi: 10.3389/fnhum.2014.00020
- Clark, A. (2012). Dreaming the whole cat: generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind* 121, 753–771. doi: 10.1093/mind/fzs106
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780190217013.001.0001
- Crick, F., and Mitchison, G. (1983). The function of dream sleep. *Nature* 304, 111–114. doi: 10.1038/304111a0
- Cutsuridis, V., and Yoshida, M. (2017). Memory processes in medial temporal lobe: experimental, theoretical and computational approaches. *Front. Syst. Neurosci.* 19. doi: 10.3389/fnsys.2017.00019
- Domhoff, G. W. (2011). The neural substrate for dreaming: is it a subsystem of the default network? *Conscious. Cogn.* 20, 1163–1174. doi: 10.1016/j.concog.2011.03.001
- Dresler, M., Eibl, L., Fischer, C. F., Wehrle, R., Spoormaker, V. I., Steiger, A., et al. (2014). Volitional components of consciousness vary across wakefulness, dreaming and lucid dreaming. *Front. Psychol.* 4, 987. doi: 10.3389/fpsyg.2013.00987
- Edwards, C., Ruby, P., Malinowski, J., Bennett, P., and Blagrove, M. (2013). Dreaming and insight. *Front. Psychol.* 4, 979. doi: 10.3389/fpsyg.2013.00979
- Fischmann, T., Ambresin, G., and Leuzinger-Bohleber, M. (2021). Dreams and Trauma Changes in the manifest dreams in psychoanalytic treatments—a psychoanalytic outcome measure. *Front. Psychol.* 12, 678440. doi: 10.3389/fpsyg.2021.678440

All authors contributed to the article and approved the submitted version.

Funding

This research discussed here is supported by the International Psychoanalytic University (IPU) Berlin, the American Psychoanalytic Association (APsaA, grant nos. FPR W20 and FPR W22), the International Psychoanalytical Association (IPA, grant nos. 3505 and 36) to TF, and the open access funding by the Publication Fund of Charité Berlin.

Acknowledgments

We would like to thank the peer reviewers for their thoughtful and constructive comments on previous versions of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Fischmann, T., and Leuzinger-Bohleber, M. (2017). Veränderungen von Träumen als Indikatoren für Therapieerfolg. *Trauma—Zeitschrift Für Psychotraumatologie Und Ihre Anwendungen* 15, 80–89.
- Fischmann, T., Russ, M. O., and Leuzinger-Bohleber, M. (2013). Trauma, dream, and psychic change in psychoanalyses: a dialog between psychoanalysis and the neurosciences. *Front. Hum. Neurosci.* 7, 877. doi: 10.3389/fnhum.2013.00877
- Fox, K. C., Nijeboer, S., Solomonova, E., Domhoff, G. W., and Christoff, K. (2013). Dreaming as mind wandering: evidence from functional neuroimaging and first-person content reports. *Front. Hum. Neurosci.* 7, 412. doi: 10.3389/fnhum.2013.00412
- Freud, S. (1894). *The Neuro-Psychoses of Defense. Standard Edition, Vol. 3*. London: Hogarth Press. p. 45–61.
- Freud, S. (1900). “The interpretation of dreams Sigmund Freud,” in J. Strachey, J. Strachey, ed (Avon: New York).
- Freud, S. (1915). “Instincts and their vicissitudes,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV (1914–1916): On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*. (p. 109–140).
- Freud, S. (1920). “Beyond the pleasure principle,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XVIII (1920–1922): Beyond the Pleasure Principle, Group Psychology and Other Works*. pp. 1–64. doi: 10.1037/11189-000
- Freud, S. (1923). *The Ego and the Id*. London: Hogarth Press and Institute of Psychoanalysis.
- Freud, S. (1940). *An outline of Psychoanalysis*. New York: Penguin
- Friston, K. (2012). Prediction, perception and agency. *Int. J. Psychophysiol.* 83, 248–252. doi: 10.1016/j.ijpsycho.2011.11.014
- Griffiths, R. R., Richards, W. A., McCann, U., and Jesse, R. (2006). Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology*, 187, 268–283. doi: 10.1007/s00213-006-0457-5
- Hobson, J. A. (2014). Consciousness, dreams, and inference: the Cartesian theater revisited. *J. Consciousness Stud.* 21, 6–32.
- Hobson, J. A., and Friston, K. J. (2012). Waking and dreaming consciousness: neurobiological and functional considerations. *Prog. Neurobiol.* 98, 82–98. doi: 10.1016/j.pneurobio.2012.05.003
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: OUP. doi: 10.1093/acprof:oso/9780199682737.001.0001
- Holze, F., Vizeli, P., Ley, L., Müller, F., Dolder, P., Stocker, M., et al. (2021). Acute dose-dependent effects of lysergic acid diethylamide in a double-blind placebo-controlled study in healthy subjects. *Neuropsychopharmacology* 46, 537–544. doi: 10.1038/s41386-020-00883-6
- Koslowski, M., Johnson, M. W., Gründer, G., and Betzler, F. (2021). Novel treatment approaches for substance use disorders: therapeutic use of psychedelics and the role of psychotherapy. *Current Addict. Rep.* 9, 1–11. doi: 10.1007/s40429-021-00401-8
- Kraehenmann, R. (2017). Dreams and psychedelics: neurophenomenological comparison and therapeutic implications. *Curr. Neuropharmacol.* 15, 1032–1042. doi: 10.2174/1573413713666170619092629
- Kraehenmann, R., Pokorny, D., Aicher, H., Preller, K. H., Pokorny, T., Bosch, O., et al. (2017). LSD increases primary process thinking via serotonin 2A receptor activation. *Front. Pharmacol.* 8, 814. doi: 10.3389/fphar.2017.00814
- Letheby, C., and Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neurosci. Consciousness* 3, doi: 10.1093/nc/nix016
- Loo, M.-R., and Cheng, S.-K. (2022). Dream lucidity positively correlates with reality monitoring. *Conscious. Cogn.* 105, 103414. doi: 10.1016/j.concog.2022.103414
- Michaël, A. M., Parker, P. R., and Niell, C. M. (2019). A hallucinogenic serotonin-2A receptor agonist reduces visual response gain and alters temporal dynamics in mouse V1. *Cell Rep.* 26, 3475–3483.e3474. doi: 10.1016/j.celrep.2019.02.104
- Millière, R. (2017). Looking for the self: phenomenology, neurophysiology and philosophical significance of drug-induced ego dissolution. *Front. Hum. Neurosci.* 11, 245. doi: 10.3389/fnhum.2017.00245
- Millière, R., Carhart-Harris, R. L., Roseman, L., Trautwein, F.-M., and Berkovich-Ohana, A. (2018). Psychedelics, meditation, and self-consciousness. *Front. Psychol.* 9, 1475. doi: 10.3389/fpsyg.2018.01475
- Moser, U., and Hortig, V. (2019). *Mikrowelt Traum—Affektregulierung und Reflexion*. Frankfurt a. M.: Brandes and Apsel.
- Moser, U., and von Zeppelin, I. (1996). *Der geträumte Traum: wie Träume entstehen und sich verändern*. Kohlhammer.
- Mota-Rolim, S. A. (2020). On moving the eyes to flag lucid dreaming. *Front. Neurosci.* 14, 361. doi: 10.3389/fnins.2020.00361
- Palhano-Fontes, F., Mota-Rolim, S., Lobão-Soares, B., Galvão-Coelho, N., Maia-Oliveira, J. P., Araújo, D. B., et al. (2021). “Recent evidence on the antidepressant effects of ayahuasca,” in *Ayahuasca Healing and Science* (Berlin: Springer). p. 21–41. doi: 10.1007/978-3-030-55688-4_2
- Panksepp, J., and Wilson, C. G. (2016). “Brain SEEKING circuitry in neuroeconomics: a unifying hypothesis for the role of dopamine-energized arousal of the medial forebrain bundle in enthusiasm-guiding decision-making,” in *Neuroeconomics* (Berlin: Springer). p. 231–252. doi: 10.1007/978-3-642-35923-1_13
- Preller, K. H., and Vollenweider, F. X. (2018). Phenomenology, structure, and dynamic of psychedelic states. *Curr. Top. Behav. Neurosci.* 36, 221–256. doi: 10.1007/978-3-642-35923-1_13
- Revonsuo, A. (2000). The reinterpretation of dreams: an evolutionary hypothesis of the function of dreaming. *Behav. Brain Sci.* 23, 877–901. doi: 10.1017/S0140525X00004015
- Robbins, M. (2018). The primary process: Freud’s profound yet neglected contribution to the psychology of consciousness. *Psychoanalytic Inquiry* 38, 186–197. doi: 10.1080/07351690.2018.1430965
- Ruby, P. M. (2011). Experimental research on dreaming: state of the art and neuropsychanalytic perspectives. *Front. Psychol.* 2, 286. doi: 10.3389/fpsyg.2011.00286
- Sanz, C., Zamberlan, F., Erowid, E., Erowid, F., and Tagliazucchi, E. (2018). The experience elicited by hallucinogens presents the highest similarity to dreaming within a large database of psychoactive substance reports. *Front. Neurosci.* 12, 7. doi: 10.3389/fnins.2018.00007
- Scarpelli, S., Nadorff, M. R., Bjorvatn, B., Chung, F., Dauvilliers, Y., Espie, C., et al. (2022). Nightmares in people with COVID-19: did Coronavirus infect our dreams? *Nat. Sci. Sleep* 14, 93. doi: 10.2147/NSS.S344299
- Schmidt, T. T., and Berkemeyer, H. (2018). The altered states database: psychometric data of altered states of consciousness. *Front. Psychol.* 9, 1028. doi: 10.3389/fpsyg.2018.01028
- Schultes, R. E., and Hofmann, A. (1979). *Plants of the Gods: Origins of Hallucinogenic Use*. New York: McGraw-Hill.
- Solms, M. (2000). Dreaming and REM sleep are controlled by different brain mechanisms. *Behav. Brain Sci.* 23, 843–850. doi: 10.1017/S0140525X00003988
- Solms, M. (2013). The conscious id. *Neuropsychanalysis* 15, 5–19. doi: 10.1080/15294145.2013.10773711
- Solms, M. (2014). *The Neuropsychology of Dreams: A Clinico-Anatomical Study*. Milton Park: Psychology Press. doi: 10.4324/9781315806440
- Solms, M. (2018). The neurobiological underpinnings of psychoanalytic theory and therapy. *Front. Behav. Neurosci.* 12, 294. doi: 10.3389/fnbeh.2018.00294
- Solms, M. (2019). The hard problem of consciousness and the free energy principle. *Front. Psychol.* 9, 2714. doi: 10.3389/fpsyg.2018.02714
- Solms, M. (2021). *The Hidden Spring: A Journey to the Source of Consciousness*. New York: WW Norton and Company. doi: 10.53765/20512201.28.11.153
- Spanò, G., Pizzamiglio, G., McCormick, C., Clark, I. A., De Felice, S., Miller, T. D., et al. (2020). Dreaming with hippocampal damage. *Elife* 9, e56211. doi: 10.7554/eLife.56211.sa2
- Spiegelhalter, D., and Rice, K. (2009). Bayesian statistics. *Scholarpedia* 4, 5230. doi: 10.4249/scholarpedia.5230
- Stigler, M., and Pokorny, D. (2001). Emotions and primary process in guided imagery psychotherapy: computerized text-analytic measures. *Psychother. Res.* 11, 415–431. doi: 10.1093/ptr/11.4.415
- Studerus, E., Komater, M., Hasler, F., and Vollenweider, F. X. (2011). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: a pooled analysis of experimental studies. *J. Psychopharmacol.* 25, 1434–1452. doi: 10.1177/0269881110382466
- Sui, J., and Humphreys, G. W. (2015). The integrative self: how self-reference integrates perception and memory. *Trends Cogn. Sci.* 19, 719–728. doi: 10.1016/j.tics.2015.08.015
- Tuominen, J., Stenberg, T., Revonsuo, A., and Valli, K. (2019). Social contents in dreams: an empirical test of the social simulation theory. *Conscious. Cogn.* 69, 133–145. doi: 10.1016/j.concog.2019.01.017
- Vollenweider, F. X., and Komater, M. (2010). The neurobiology of psychedelic drugs: implications for the treatment of mood disorders. *Nat. Rev. Neurosci.* 11, 642–651. doi: 10.1038/nrn2884
- Windt, J. M. (2015). *Dreaming: A Conceptual Framework for Philosophy of Mind and Empirical Research*. Cambridge: MIT press. doi: 10.7551/mitpress/9780262028677.001.0001
- Wittmann, L., Anstadt, T., Fischmann, T., Hau, S., Kempe, S., Herot, K., et al. (2017). Ein Traum, zwei Methoden. *J. Psychoanalyse* 58, 99–129. doi: 10.18754/jfp.58.6
- Wolff, M., Evens, R., Mertens, L. J., Koslowski, M., Betzler, F., Grunder, G., et al. (2020). Learning to let go: a cognitive-behavioral model of how psychedelic therapy promotes acceptance. *Front Psychiatry* 11, 5. doi: 10.3389/fpsyg.2020.00005
- Yaden, D. B., Earp, D., Graziosi, M., Friedman-Wheeler, D., Luoma, J. B., Johnson, M. W., et al. (2022). Psychedelics and psychotherapy: cognitive-behavioral approaches as default. *Front. Psychol.* 13, 1604. doi: 10.3389/fpsyg.2022.873279



OPEN ACCESS

EDITED BY

Filippo Cieri,
Cleveland Clinic, United States

REVIEWED BY

Eamonn Walsh,
King's College London, United Kingdom
Marcus Heldmann,
University of Lübeck, Germany

*CORRESPONDENCE

Lola Thieffry
✉ lola.thieffry@ulb.be

SPECIALTY SECTION

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

RECEIVED 31 August 2022

ACCEPTED 03 February 2023

PUBLISHED 29 March 2023

CITATION

Thieffry L, Olyff G, Pioda L, Detandt S and
Bazan A (2023) Running away from
phonological ambiguity, we stumble upon our
words: Laboratory induced slips show
differences between highly and lowly
defensive people.
Front. Hum. Neurosci. 17:1033671.
doi: 10.3389/fnhum.2023.1033671

COPYRIGHT

© 2023 Thieffry, Olyff, Pioda, Detandt and
Bazan. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Running away from phonological ambiguity, we stumble upon our words: Laboratory induced slips show differences between highly and lowly defensive people

Lola Thieffry^{1,2*}, Giulia Olyff², Lea Pioda^{3,4}, Sandrine Detandt^{2,4}
and Ariane Bazan^{1,2,5}

¹Laboratoire InterPsy (UR 4432), Université de Lorraine, Nancy, France, ²Observatoire du Sida et des Sexualités, Université Libre de Bruxelles, Brussels, Belgium, ³Parhélie Asbl, Institution Psychiatrique, Brussels, Belgium, ⁴Faculté des Sciences Psychologiques et de l'Éducation, Université Libre de Bruxelles, Brussels, Belgium, ⁵Centre de Recherche en Psychologie Clinique, Psychopathologie et Psychosomatique, Université Libre de Bruxelles, Brussels, Belgium

Introduction: Freud proposed that slips of the tongue, including apparently simple ones, always have a sense and constitute « a half-success and a half-failure » compromise resulting from defensive mechanisms.

Material and methods: A total of 55 subjects participated in a French adaptation of the *Spoonerisms of Laboratory Induced Predisposition* or SLIP-technique including 32 “neutral” and 32 taboo spoonerisms and measures of defensiveness. In accordance with a psychoanalytical and empirically supported distinction, we considered two kinds of defenses: elaborative or primary process and inhibitory or secondary process defenses, which were operationalized with the GeoCat and the Phonological-Nothing (PN) WordList, respectively. The GeoCat is a validated measure of primary process mentation and the PN WordList was shown to measure the defensive avoidance of language ambiguity.

Results: Participants produced 37 slips, with no significant difference in the number of “neutral” and taboo slips. The GeoCat and the N/PN parameters explained 30% of the variance in the production of parapraxes, confirming the defensive logics of slips. When dividing the population into lowly and highly defensive participants (with the Marlowe Crowne Social Desirability scale), primary process mentation appears as a baseline default defense, but only highly defensive participants mobilize an additional inhibitory secondary process type of defense. Taking into account the *a priori* difference between taboo and “neutral” parapraxes, highly defensive participants made 2.7 times more taboo parapraxes than lowly defensive participants. However, if “neutral” parapraxes in both subgroups followed the same logic as the total group of parapraxes (significant contribution of primary process mentation in lowly defensives and of primary and secondary process mentation in highly defensives), these measures had no contribution to explain the occurrence of taboo parapraxes.

Conclusion: We propose that Motley et al.'s prearticulatory editor, ensuring the censorship over taboo parapraxes, is an external instance of inhibition, proximal to uttering, equivalent to the censorship between the systems Preconscious and Conscious in Freud's metapsychology. By contrast, the

defenses measured in this research are internal, intimate control systems, probing for the censorship between the systems Unconscious and Preconscious, this is, for repression. This study contributes to support a psychodynamic explanatory model for the production of parapraxes.

KEYWORDS

Freud, slips of the tongue, repression, defense, metapsychology, unconscious, preconscious, signifier

1. Introduction

A little ad in *The Andover Townsman*, the hometown newspaper of the American city of Andover, on January 3, 1913 (p. 7) goes as follows: « Mr. Spooner (...) was very shy and would never have had the courage to ask a woman to be his wife, but one afternoon, in a friend's drawing room, he was requested to ask one of the ladies present to make tea. In doing so, he blundered as usual: "Will you take me?" he said, instead of "Will you make tea?." Blushing, the lady "took him" and thus he "blundered" into a happy marriage ». William Archibald Spooner (1844–1930) was an English clergyman, known for making spoonerisms frequently, so much that the word "spoonerism" is in fact derived from his last name. Spoonerisms (in French, *contrepèteries*) typically occur when transposing corresponding sounds or phonemes in words and are to be considered as a special case of slips of the tongue (Motley and Baars, 1976a). Slips of the tongue, or parapraxes, are an intriguing behavioral phenomenon dividing the field of psychology. In *Psychopathology of Everyday Life*, Freud (1901/1978, p. 271) assumes that daily life failures, such as forgetting names and words, verbal parapraxes, reading and writing errors, are not fortuitous but « have a hidden motivation ». A conflictual dynamic lies at the heart of these parapraxes which is, at least by some aspects, unacceptable to conscious thought and against which we defend ourselves. However, this defense only succeeds incompletely and will therefore give itself away (partially). For example, Freud (1901/1978, pp. 276–277, p. 279, *Italics added*) states that they bear « to the greater or lesser degree (...) the marked character of being "repressed" » and « can be traced back to *incompletely suppressed psychical material*, which, although pushed away by consciousness, has nevertheless not been robbed of all capacity for expressing itself ». This interpretation of verbal slips has become quite popular in the general public, so much so that it has given rise to the French expression *lapsus révélateurs* or "revelatory parapraxes."

At the metapsychological level, slips of the tongue appear as an opportunity for the "system Unconscious (Ucs)" to transcend the two censors of the first topic, i.e., to get invested by the preconscious and thereupon to find an exceptional direct access to the "system Conscious (Cs)" and to voluntary motility, leading the subject to suddenly get a direct hear of his proper unbearable thoughts, and often creating embarrassment. However, psycholinguists both in Freud's time (e.g., Meringer and Mayer, 1895) and nowadays (e.g., Rossi and Peter-Defare, 1998) consider parapraxes as accidental speech errors caused by linguistic and cognitive mechanisms and dismiss the idea that slips could be caused by repressed thoughts. Indeed, psycholinguistically, such errors are seen as failures of

error control systems: while the selection of words progresses in spoken language production, either self-monitoring systems control systems (in Levelt's serial model; e.g., Levelt, 1989; Levelt et al., 1999) or feedback dynamics (in Dell and Reich's connectionist model; Dell and Reich, 1980, 1981) are supposed to correct for possible production errors. Moreover, in the connectionist model (Dell and Reich, 1980, 1981) word selection is subject to influences through associative priming by preceding words and to neighborhood activation mechanisms. In fact, this way of considering slips is not too far from Freud's line of thought when he says: «... the positive factor favoring the slip of the tongue (the uninhibited stream of associations) and the negative factor (the relaxation of the inhibiting attention) invariably achieve their effect in combination, so that the two factors become merely different ways of regarding the same process. What happens is that, with the relaxation of the inhibiting attention—in still plainer terms, as a result of this relaxation—the uninhibited stream of associations comes into action » (Freud, 1901/1978, p. 61). The "uninhibited stream of associations" might be seen as quite equivalent to the spreading activation in a connectionist model and psychoanalytically amounts to primary process mentation (see further), while the "inhibiting attention" might be seen as equivalent to either the retroactive feedback mechanisms in the connectionist model or as the self-monitoring module in the serial language production model, and psychoanalytically as secondary process mentation (see further). Conceived as such, "the two factors" (spreading activation and inhibition) can be conceived of as the complementary activation of primary and secondary process mentation.

Despite these commonalities between the psycholinguistic and the psychodynamic model, deep divergences remain. Indeed, Freud (1916–1917/1966, p. 44) is keen to specify that parapraxes are not the result of a subject-less mechanism: « They are not chance events but serious mental acts; they have a sense ». In most cognitive models, the basic stages of language production, including the phonological preparation, are considered as automatic stages (e.g., Levelt, 1989), i.e., as subject-less (« without the necessity for active control or attention by the subject »; Schneider and Shiffrin, 1977, p. 2). However, Freud obliges us to suppose an intentional subject already at the level of the "basic" activation and inhibition-mechanisms underlying the production of language. This influence of the subjective structure—most patently, anxiety, and defensiveness—indeed explained results in subliminal linguistic priming research (Klein Villa et al., 2006; Bazan et al., 2019a): only anxious subjects showed subliminal palindrome priming and only defensive subjects showed subliminal aversion for phonological

ambiguity. According to Shevrin (1992) personality factors do not influence cognitive tasks at the conscious level but could do so at an unconscious level.

This is also what results from psycholinguistic research on parapraxes. Indeed, verbal parapraxes can be induced experimentally with the *Spoonerisms of Laboratory Induced Predisposition* or “SLIP”-technique (Motley and Baars, 1976a). The SLIP-task asks participants to silently read word couples with shared phonemes, thereby sometimes reading out loud certain target-cued pairs. These target pairs, for example, “balm peach,” are designed to produce spoonerisms such as: *balm peach* → *palm beach* and are therefore spoonerism eliciting pairs (from here simply called “eliciting pairs”). This means that by interchanging the initial consonants, a new pair of word emerges which differs in meaning with the target pair. Typically, eliciting pairs are preceded by several phonological interference word pairs that resemble the phonology of the expected spoonerism and increase the probability of producing a slip of the tongue (Motley and Baars, 1976a). For example, *barred dorm* and *bought dog* precede the target *darn bore* that erroneously could be uttered as *barn door*. Motley (1985) claimed that these artificial slips of the tongue are similar to those occurring naturally. Motley et al. (1981a, 1982) also have demonstrated that participants intercept taboo errors more often than neutral ones proposing that an « automatic » and « subconscious » (Motley et al., 1979, p. 196) prearticulatory component of speech production “censors” the overt formulation of taboo spoonerisms because of their socially inappropriate character. Interestingly, it has led these psycholinguistic authors, as well as others after them (e.g., Severens et al., 2011, 2012; Wagner-Altendorf et al., 2020), to make an intuitive distinction between “taboo” (e.g., *tool kits* → *cool tits*) and “neutral” (e.g., *darn bore* → *barn door*) parapraxes.

A taboo designates « something that is not acceptable to say, mention, or do on grounds of morality or taste » (Webster’s Dictionary). Taboo words designate more widely themes having to do with sexuality, death, racism, bodily productions, insults, etc. (Jay et al., 2008) and are distinct from words that “simply” have negative emotional valence (Jay, 1999; Hansen et al., 2019). During childhood, taboo words are recognized as such, through education and socialization, their use being repressed by care and authority figures (Jay, 2009). Independently of the SLIP-methodology, a number of psycholinguistic studies including taboo words (Hartsuiker and Kolk, 2001; Dhooge and Hartsuiker, 2011; White et al., 2017) have corroborated the existence of a “verbal self-monitor” which would both identify and correct errors during the speech production process and thwart the highly embarrassing utterance of taboo words, simultaneously slowing down their response times.

The Motley et al. (1981a, 1982) finding of fewer taboo than neutral slips, has thus been interpreted as the result of a prearticulatory editor, which censors the taboo slips before they are uttered. This suggests that neutral slips, in contrast, would be more likely the result of failures in the word production system. However, this is at odds with Freud, (1901/1978, p. 83), who is quick to underscore that « *even apparently simple slips* of the tongue could be traced to interference by a half-suppressed idea that lies outside the intended context » (Freud, 1901/1978, p. 83, Italics added) and: « In contrast to these groups of cases, in which the parapraxis itself brings its sense to light, there are others in which the parapraxis

produces nothing that has any sense of its own, and which therefore sharply contradict our expectations. If someone twists a proper name about by a slip of the tongue or puts an abnormal series of sounds together, these very common events alone seem to give a negative reply to our question whether all parapraxes have some sort of sense. Closer examination of such instances, however, shows that these distortions are easily understood and that « *there is by no means so great a distinction between these more obscure cases and the earlier straightforward ones* » (Freud, 1916–1917/1966, pp. 41–42, Italics added). In other words, these “neutral” slips are not to be considered as “system glitches”; as a matter of fact, their target (the slip) would be precisely aimed at. It is clear that for Freud there are no *a priori* distinctions between taboo and so-called “neutral” parapraxes. Indeed, when we examine specific examples of speech blunders, they often include errors on words which acquire an emotional significance in the given context, but which outside this context would undoubtedly be deemed “neutral”. For example, in Motley et al. (1982), the potential spoonerism *darn bore* → *barn door* is said neutral. However, Klein Villa et al. (2006) recount the following anecdote: an audience member at a conference who intends to communicate that he would like the door closed in order to better hear the speaker, but is simultaneously distracted due to an inner state of boredom, states « close the bore » instead of « close the door ». We see here that this slip reveals the transgression of a taboo, namely, insulting someone. For all these reasons, there is no *a priori* ground to reject the idea that every parapraxis also has a defensive component. In summary, our first hypothesis is that **(1) the production of parapraxes is always a defensive process, both for so-called neutral and taboo-parapraxes.**

A specific difficulty is the measurement of defensive processes. We propose that, in line with what Freud (1895/1966) describes in his *Project*, mental processes are defensive by essence: this is their ontological principle. The very reason why a mental system arises is to ward off accumulating stimulation, which threatens to burn (the membranes of) the organism; moreover, a mental apparatus grows in complexity in order to simultaneously directly discharge excess excitation and retain a fraction for the more elaborate execution of specific actions. The primary process entails the flight for incoming excitations by the shortest pathway possible: this neuronal dynamic organization is reflected in associative thinking, ruled by the pleasure principle and the overall outcome is a search for perceptual identity (Freud, 1900/1955). This perceptual identity involves the recognition and the identification of similar or identical elements that have only a fragment or attribute in common—that is, « superficial » (Freud, 1900/1955, p. 597) or « non-essential » (Holt, 1967, p. 334) similarities. However, when it comes to our internal needs (e.g., hunger) fleeing the stimulus is of no avail, and the « removal of the stimulus is only made possible here by an intervention which (...) calls for an alteration in the external world (supply of nourishment, proximity of the sexual object), which, as a specific action, can only be brought about in definite ways. » (Freud, 1895/1966, p. 316) in order to ward off this threatening tension. For this to happen, the organism « must put up with (maintaining) a store of Q (quantities of excitation) sufficient to meet the demand for a specific action » (Freud, 1895/1966, p. 297). This is the secondary process, which also inhibits primary process associative reactions (Freud, 1895/1966). In other words, a human mental system develops both primary process defense, the direct mirror-like discharge and secondary process defense, a more

organized discharge also involving the inhibition of the primary process discharge. In this regard, primary and secondary process dynamics are simultaneously constitutive and defensive, be it with another functional principle (see also Bazan, submitted¹).

This distinction is parallel to Erdelyi's (2006) proposition who, in his *unified theory of repression*, also divides repression into two subclasses, either additive and elaborative or subtractive and inhibitory. « In inhibitory repression », Erdelyi (2006, p. 502) says, « the consciousness-lowering operation is readily conceived of as some type of psychological subtraction that results in lower consciousness (e.g., we subtract attentional allocation from a channel, we reduce or eliminate thinking about some material) ». He underscores that Freud's initial conception of repression was of the inhibitory or subtractive variety. Freud, (1900/1955, p. 599, p. 601) literally says: the secondary process « succeeds in *inhibiting* this discharge (from the primary process) » and refers to « the inhibition imposed by the second system as the “secondary process” ». Saraga and Gasser (2005, p. 111) indicate that Freud underscored the importance of this inhibition as being the essence of the secondary process. Scano (2007, p. 141) says: « defense (...) functions by inhibiting the primary process and progressively establishing the secondary process ». For these reasons, the secondary process, which is an inhibitory type of defense, is of the kind which Erdelyi had in mind, this is, in essence, repression (see also further). Indeed, we have argued elsewhere that repression is a special instance of inhibition for highly invested *linguistic* stimuli (Bazan, 2012, p. 13). Articulation is also a « specific action » (see Bazan, 2007). For example, d'Épinay (2003, p. 88, our translation) says: « In the Project, Freud considers that this “innervation of speech” is originally a discharge mechanism, a safety valve ensuring a temporary and partial decrease in tension, along non-specific pathways, until the discovery of the “specific action” ». Specifically, this inhibition is possibly instantiated physiologically by the efference-copy-mediated attenuation of predictable proprioceptive return of the *articulation of the linguistically* grasp on stimuli (see also Bazan and Snodgrass, 2012): briefly, when initiating an action of will we predict the sensorimotor repercussions on the proper body of that action (the new positions of the muscles, the joints and the skin once the command will be carried out) and on the basis of this prediction, by anticipation, attenuate the sensorimotor cortices to that predicted level so that when the stimulation indeed comes, the feeling is readily neutralized. We have proposed that this mechanism is also the mechanism by which we attenuate the hearing of innuendos, ambiguities, and peculiar associations (see also Haskell, 1991, 2001),² i.e., that

it contributes to the mechanism of repression (Bazan, 2012). Erdelyi (2006) further underscores that the inhibitory-type of repression results in rebound phenomena, this is the return-of-the-repressed. In our linguistic model of repression (Bazan, 2012), unconscious inhibition of the specific articulatory action would induce return of the same *articulatory* fragments—mostly, however, as radically different meanings (e.g., the “rat obsession” in the Ratman, etiologically linked to *Heiraten* and *Frau Hofrat*), which is, of course, a radically efficient way of masking (Bazan, 2007, 2012), and thus of circumventing censorship. Specifically in parapraxes, the slips of the tongue appear as ideal ways for the return-of-the-repressed after an inhibitory kind of defense, i.e., after repression. For all these reasons, repression and the secondary process—involving the specific act of speaking and its inhibition—are logically equivalent, and of the inhibitory defense type.

For the elaborative kind of defenses, Erdelyi (2006) quotes rationalization, projection, reversal, displacement, and symbolization. Displacement is undoubtedly a primary mental process (Freud, 1900/1955) and so are symbolization (Rapaport, 1951, p. 694) and projection (Rapaport, 1951, p. 690). But for reversal too, this is obviously the case: « reversal, or *turning a thing into its opposite*, is one of the means of representation most favored by the dream-work (...) » (Freud, 1900/1955, p. 327, *Italics added*). In other words, reversal appears as a mirror-wise equation of one line of thought by another, which is a typical primary process fashion of mental processing. And this is true for rationalization, too; e.g., « Here is an example of (...) [an] attempt to derive one symptom from another by means of an intellectual rationalization: it is suggested that the patient, who, owing to a primary disposition, believes that he is being persecuted, infers from this persecution that he must be someone of quite particular importance and so develops megalomania » (Freud, 1916–1917/1966, p. 424). But rationalization as a defense mechanism is not limited to delusions: Freud understands rationalization as an operation that fulfills functions in the mental life serving the pleasure principle and independently of its degree of truth (e.g., Freud, 1914/1964, p. 52; Freud, 1933/1966, p. 542). Obviously, this puts rationalization under the banner of primary process defense mechanisms. Undoubtedly, then, the defense mechanisms, characterized by Erdelyi as the “elaborative kind,” function on the primary process mode. Indeed, in schizophrenia we observe primary process mentation as a defense in a straightforward way when, upon having to deal with an unexpected, and therefore “intrusive,” stimulation, a subject decompensates in delusional attribution of meaning. Our own patient Hervé (Bazan, 2012, p. 5) develops associative train thoughts on a primary process mode when intruded by visual stimuli he could not predict.

For all these reasons, we support Erdelyi's dichotomous division of defense mechanisms in either elaborative or inhibitory dynamics and understand it as the division in primary and secondary process defense mechanisms. When it comes to parapraxes specifically, we are backed up by Freud, (1901/1978, p. 61), who proposes that parapraxes can be the result both of associative speech production

1 Bazan, A. (submitted). Primary and secondary process mentation and the pertinence of the Freudian model for modern science.

2 Language is supposed to be full of such innuendo's (see also Freud, 1901/1978, p. 239) but (because of repression), we mostly don't hear them. The psychoanalyst Gauthier Lafaye tells a story about one of his patients whose father had left the house when she was a kid. During a session, the woman tells « Ma mère n'était pas parvenue » (« My mother did not succeed in ») but pauses when she says « par-venue », leading the analyst to hear another meaning « papa revenue » (« daddy has come back »). When the analyst simply repeats « papa r'venu » this opens a new line of emotional associations, leading the patient to express, for the first time, her grief about the loss of her father. Independently from psychoanalysis, the psychologist Robert Haskell has proposed that the thickness of enunciation in a social exchange also includes a layer which is not consciously experienced, which is articulated at the level of phonological polysemy and which deals with the relationship between the speakers. He has even developed

a methodology to track such dual meanings in speech (Haskell, 2003). His matrix proposes in a particular exchange, for example, the phonemic reading of the word group « The stuffy nose » as « The stuff he knows », and together with other example, he suggests that even in apparently innocuous descriptions, peoples express their (unconscious) preoccupations with their mutual relationships of power and seduction.

(« the uninhibited stream of associations » or “positive” primary process defense) but also as the return of speech fragments which were previously put under inhibition, i.e., under tension—in other words as the return-of-the-repressed (« the relaxation of the inhibiting attention »). If we agree upon this perspective, we may now propose measures for primary process mentation and for secondary process inhibition, respectively.

For primary process mentation, the Geometrical Categorization Task (or GeoCat; Brakel et al., 2000) is a validated measure (for review see Bazan and Brakel, 2023). The GeoCat is a simple, non-verbal tool which maps preferential mobilization of primary versus secondary processes in the treatment of mental stimuli by asking participants to make similarity judgments between geometrical figures. The theoretical background is a cognitive theory of categorization that distinguishes between attributional (ATT) and configurational (REL for “relational”) similarity judgments (Smith and Medin, 1981; Murphy and Medin, 1985; Medin et al., 1990). Attributional similarity refers to the superficial resemblance between attributes of the stimuli, which is indeed the associative logic of the primary process. The specific type of relational similarity in the GeoCat is configurational with the same spatial arrangement of the components of both stimuli. Indeed, the secondary process, thanks to the “store of excitations,” which constitutes a third point, enables perspective-taking, giving access to spatiotemporal distinctions (Bazan, 2007; see text footnote 1) and this fits well with the identification of configurational similarity between stimuli in the REL items. Hence, primary process mentation is thought to be probed by the number of “attributional choices” (or ATT) and secondary process mentation by the number of “relational choices” (or REL).

As concerns the measure of the secondary process, as the GeoCat asks for forced choices between ATT and REL, the REL-choices are not independent secondary process measures. The GeoCat has in the past proven especially interesting to probe for primary process mentation (e.g., Bazan et al., 2013, 2019b). But primary and secondary process dynamics are not mutually exclusive, since they operate conjointly (e.g., Green, 1995) and are supposed to concur into the production of parapraxes. Therefore, an independent measure, specifically targeted upon linguistic inhibition, is needed here. In two previous instances in our research, we have productively made use of such a measure, namely the so-called “Phonological-Nothing WordList.” In the PN WordList, participants are presented with a prime word (e.g., *Nice* /-nais/) and are asked to make a similarity forced-choice between a Phonological (P) target (namely, a phonological inverse, e.g., *Sign* /-sain/) and a Non-related (N) related target (e.g., *Belt*). Previous results have shown that choosing N in PN can also be understood as a *negative* choice for P: participants were thought to choose P (also) because they try to escape from phonological ambiguity (Bazan et al., 2019a). Thus, this negative choice for P would show a defensive move indicating an aversion for phonological ambiguity. This was confirmed in subliminal presentation by two independent measures of social desirability, the Marlowe Crowne Social Desirability scale (MCSD) and Balanced Inventory of Desirable Responding (interpreted as measures of defensiveness, see further), which predicted both this avoidance from phonological ambiguity on the WordList ($r = -0.51$; $p = 0.004$) as well as a neural evoked potential response, the PMN

or Phonological Mismatch Negativity, interpreted as « perplexity in the face of phonological ambiguity » (Bazan et al., 2019a, p. 11). In recent results (Olyff and Bazan, 2023), a relative preference for N in PN predicts worse supraliminal rebus resolution scores. We interpreted the lesser rebus resolution in high N-PN subjects not merely as the result of a lesser phonological awareness or appetite (there were not more P-choices in a parallel Phonological-Semantic WordList), but as the result of a defensive avoidance of the ambiguous P-version of the target. For all these reasons, if the production of parapraxes is a defensive process, the number of ATT-choices in the GeoCat, as well as the number of N choices in the PN WordList made by the subject, should significantly predict the number of parapraxes produced by the subject.

Going back to Shevrin’s (1992) idea that personality plays a role at the level of operations which are deemed “automatic,” and therefore “mindless,” at a cognitive level, our second hypothesis is that personality, and more specifically, defensiveness will play a role in the production of parapraxes. Indeed, even if we think that in neurosis repression is the basic defense mechanism, we still think that neurotic persons can be lowly or highly defensive, with the level of defensiveness as a psychodynamic functioning mode, or in other terms, as a “character trait”—even if not necessarily a stable one (since people can change, e.g., through an analysis). Based upon clinical (Marin, 2011; Marmursztejn, 2013) and empirical data (Bazan et al., 2019a; Olyff and Bazan, 2023), we suppose that highly defensive people will mobilize stronger defenses, and that they will especially mobilize stronger defenses against ambiguous linguistic materials. For these reasons, we predict that **(2) highly defensive people will produce more parapraxes than lowly defensive people.** As for the operationalization of this hypothesis, defensiveness is measured with the Marlowe-Crowne Social Desirability scale (MCSD). The initial intent of this scale was to measure social desirability, i.e., the need to present oneself in a socially desirable way (Crowne and Marlowe, 1960). Typical items ask participants to respond to behaviors that are « culturally sanctioned and approved but which are improbable of occurrence » (Crowne and Marlowe, 1960, p. 350) as for example « There have been occasions when I took advantage of someone ». It assesses to what magnitude subjects can accept undesirable but nevertheless universal and undeniable truths about themselves. However, the authors considered that their scale construct also measured defensiveness (Crowne and Marlowe, 1964, p. 206) as a « personality variable in its own right ». Indeed, « such favorably biased self-appraisal [has] to entail vulnerability in self-esteem and the use of repressive defenses » (Crowne and Marlowe, 1964, p. 206). The MCSD has since been widely used to assess defensiveness (see for example Weinberger et al., 1979; Weinberger, 1990; Eysenck and Van Berkum, 1992; Mann and James, 1998; Bazan et al., 2019a). Given that clinically, high defensiveness is especially identified through restrictive, inhibited behavior and the « use of repressive defenses » (Crowne and Marlowe, 1964, p. 206), and given our previous results in subliminal research showing a high correlation between subliminal N/PN and MCSD, we are furthermore inclined to predict that **(3) highly defensive people will mobilize especially inhibitory-type defenses for the production of parapraxes.**

Finally, in line with previous research with the SLIP-task, we will also **(4) explore if the a priori difference between “neutral” and taboo parapraxes is relevant in the present study.** First, are the empirical data confirming the validity of this

a priori distinction? But furthermore, and independently of this confirmation, will highly defensive people produce more taboo parapraxes specifically? Will inhibitory defense parameters better predict taboo than “neutral” parapraxes?

In summary, our main interest for the present research is to show and start to unravel the defense mechanisms underlying the production of parapraxes.

2. Materials and methods

2.1. Participants

2.1.1. Main study

A total of 55 psychology students from the *Université libre de Bruxelles* took part in the study and received course credit for participation. They had a mean age of 20 (range 18–26, $SD = 1.8$); 47 were women and one participant declared a non-binary gender identity. All participants were fluent French speakers and 40% of them were multilingual. None of these demographic variables significantly explained the production of slips.

2.1.2. Variable evaluation studies

A total of 120, resp. 958 participants, recruited by advertisements on social media, took part in two online independent studies, on item tabooess (the degree to which the participants evaluated the word pair as “taboo”), resp. MCSD evaluation. They had a mean age of 32.55 (range 17–79, $SD = 12.5$), resp. of 27.52 (range 18–90, $SD = 9.3$). 81.1%, resp. 83.6%, of the sample were women and 1.1%, resp. 1.8%, declared a non-binary gender identity. A total of 43.9%, resp. 45.5%, of the participants were multilingual.

2.2. Materials

2.2.1. The French SLIP-task

We implemented and presented Motley et al. (1981a, 1982) SLIP-protocol using PsychoPy open-source software (Peirce et al., 2019). A total of 64 French spoonerisms were constructed: 32 taboo [e.g., *seau bain* → *beau sein* (*bat bucket* → *nice boob*)] and 32 neutral [e.g., *mauve phare* → *fauve mare* (*purple lighthouse* → *wildcat pond*)]. The “tabooness” of both word pairs implied in the spoonerism—the eliciting pair, and the spoonerism itself—were evaluated by 120 participants on a 7-point Likert scale (1 = not taboo - 7 = extremely taboo) in an independent online questionnaire study. The taboo eliciting pairs were more taboo (tabooness = 1.5 ± 0.9) than their neutral counterparts (tabooness = 1.3 ± 0.7 ; Student's *t*-test; $p < 0.001$), and the actual taboo spoonerisms were also more taboo (tabooness = 3.3 ± 1.5) than the neutral spoonerisms (tabooness = 1.3 ± 0.8 ; Student's *t*-test; $p < 0.001$). Moreover, the taboo spoonerisms were more taboo than their corresponding eliciting pairs (Student's *t*-test; $p < 0.001$), but this is not the case for the neutral spoonerisms and their corresponding eliciting pairs (Student's *t*-test; $p = 0.285$). Each target word pair was preceded by three phonological prime pairs constructed using Motley et al. algorithm (see Motley et al., 1982). These prime pairs are phonologically closely similar to the intended

slips [e.g., *faune mâche* (*fauna salad*) for the spoonerism *fauve mare* (*wildcat pond*)].

An experimental trial thus consisted of three phonological primes and a target-cued or spoonerism eliciting pair (see also Figure 1). Overall, there were 64 experimental trials (32 neutral and 32 taboo) and a random number of 3–6 fillers [e.g., *rond clair* (*light circle*)], unrelated to the experimental priming of the spoonerisms, which sometimes required a verbal response as in Motley and Baars (1976a). These filler pairs were presented in order to avoid predictability of the sound signal. All word pairs were presented for 800 ms in white against a black font with 50 ms interstimulus intervals (fixation crosses). The speech prompt was a sound signal that occurred 270 ms after the target presentation and participants had 2,500 ms to give a verbal response. All trials were randomized for each participant. Participants' responses were digitally recorded using the computer's in-built microphone. The audio files were later listened to independently by two experimenters to check for the occurrence of spoonerisms (no differences were recorded between both judges).

In accordance with the original Motley-studies (e.g., Motley and Baars, 1976b; Motley et al., 1981b) as well as with e.g., Möller et al. (2007), Costa et al. (2006), and Hartsuiker et al. (2006), spoonerisms were counted as such when at least one phoneme was exchanged (e.g., *mad bug* → *mad mug* or *mad bug* → *bad bug*). Spoonerisms thus involve both partial spoonerisms (i.e., when only one phoneme was exchanged: e.g., *mad bug* → *mad mug*) and complete spoonerisms (i.e., full exchanges: e.g., *mad bug* → *bad mug*). Verbalization errors were counted when participants gave responses that were unrelated to the priming manipulation (e.g., *mad bug* → *rad bug*). Omissions were counted when participants gave no verbal responses to the target pairs.

2.2.2. The geometrical categorization task

The GeoCat 1.3 (Brakel et al., 2000; Bazan and Brakel, 2023) is a non-verbal tool which maps preferential mobilization of primary versus secondary processes in terms of similarity between geometrical figures (see Figure 2). Each GeoCat contains 6 items composed of a master figure and two target figures. The participant has to choose the target figure that he considers the most similar to the master figure. There are 4 versions (1A, 1B, 2A, 2B); these versions were randomly attributed to the participants. Series 1A and 1B are identical except that the two lower target figures in the triads are reversed (left-right). The same is true for the series 2A and 2B. This controls the possible effects of lateralization in the target choice. There were no significant differences between the 4 versions in participants' responses (Kruskal-Wallis $H = 3.58$; $p = 0.312$). The internal consistency of the GeoCat was investigated with Cronbach's Alpha ($\alpha = 0.71$).

2.2.3. The PN WordList

The PN WordList (see e.g., Bazan et al., 2019a) consists of 20 French word triads presenting a prime word [e.g., *Note* (*Grade*)] together with a phonologically related target P [e.g., *Tonne* (*Ton*)], which was actually a phonological inverse, and a neutral target N that has no intended relationship, phonological nor semantical, to the prime word [e.g., *Barbe* (*Beard*)]. The participants are asked to choose which of the two target words they think is most similar to the master word. All word triads were randomized for each

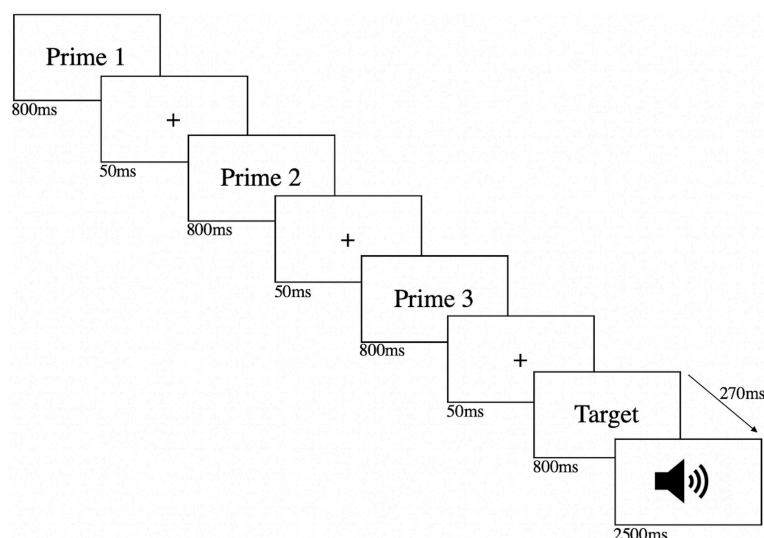


FIGURE 1

Schematic illustration of an experimental trial in the SLIP-task. Participants are shown three phonological primes before being presented with a target-cued pair (i.e., the spoonerism eliciting-pair) they have to read out loud.

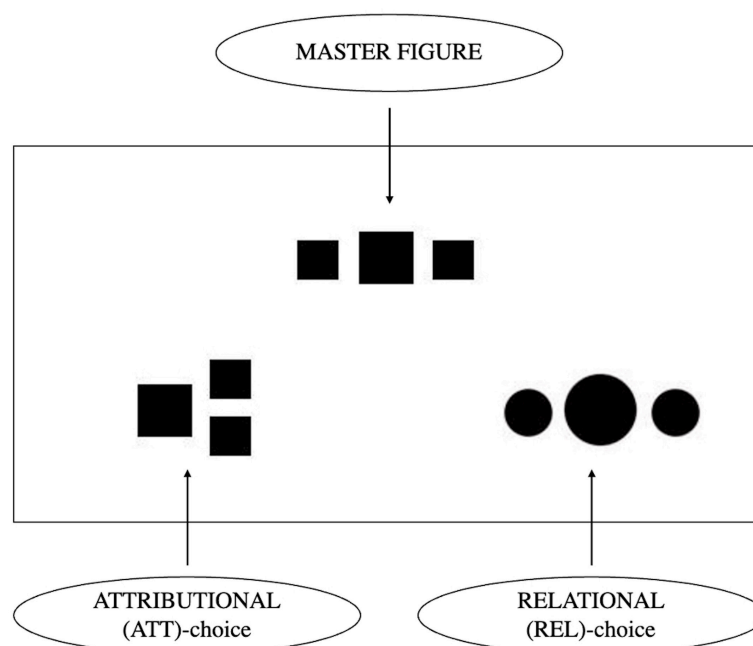


FIGURE 2

Example of a triad of one item of the Geometrical Categorization Task consisting of a master figure and two target figures. The ATT-target (**left**) consists of the same components as the master figure but in a different configuration and is thought to probe for primary process mentation; the REL-target (**right**) is made up of different components, but these are arranged in the same total configuration and is thought to probe for secondary process mentation.

participant. The internal consistency, investigated with Cronbach's α , was 0.86.

2.2.4. The Marlowe-Crowne Social Desirability scale

The MCSD (Crowne and Marlowe, 1960; French translation by Vézina, 1989) is a 33-item true/false self-report questionnaire.

Given the fact that the MCSD has suffered validity criticisms (see e.g., Leite and Beretvas, 2005), we have investigated its convergent validity in a separate, independent study ($N = 958$) and found significant correlations with a validated French social desirability measure, the DS-36 (Tournois et al., 2000; $r = 0.67$; $p < 0.001$) as well as with the "repressive defensiveness" subscale of the validated Weinberger Adjustment Inventory or WAI [French translation by

Paget et al. (2010); $r = 0.60$; $p < 0.001$]. We also investigated its divergent validity with the trait-subscale of the State Trait Anxiety Inventory [Spielberger et al., 1970; French translation by Bruchon-Schweitzer and Paulhan (1990)] yielding again predicted results ($r = -0.32$; $p < 0.001$). The internal consistency Cronbach's Alpha of the SLIP-MCSD and of the evaluation study-MCSD were 0.49 and 0.75, respectively.

2.2.5. Procedure

The experiments were conducted at the *Université libre de Bruxelles* (ULB). After a short introduction to the laboratory, participants signed an informed consent statement. Participants then were seated in front of a computer screen and received oral instructions. The experimenter was sitting behind so that he was outside of the participants' field vision. First, they had to complete a brief demographic questionnaire (age, gender, level of French, spoken languages) as well as the GeoCat 1.3. Participants were then invited to do the SLIP-task that actually was presented as a memory experiment for which they were asked to memorize each word pair for a later recall test. This strategy increases the probability of producing spoonerisms (Motley and Baars, 1976a). Our instruction was « *The aim of the present task is to study the memorization of word sequences. For this purpose, word pairs will appear quickly and successively on the screen. Try to remember as many as possible, a memory task will be presented at the end of this task. Please pay attention: some word pairs will be followed by a sound signal; this indicates that you must say aloud the last pair you saw on the screen. Pairs that are not followed by a sound signal should be read silently, internally* ». Immediately after the SLIP-task, participants had 10 min to complete a fake memory test where they were asked to write down all the pairs they had remembered.³ Next, they were invited to complete a series of questionnaires including the PN WordList and the MCSD.

3. Results

3.1. Behavioral results

Participants produced 37 spoonerisms (27 partial and 10 complete), 162 verbalization errors, and 37 omissions: on a total of 3520 trials this corresponds to, respectively, 1% spoonerisms, 4.6% verbalization errors, and 1% omissions. In regard to the number of participants, 22 made spoonerisms, 42 made verbalization errors, and 12 made omissions; on a total of 55 participants this corresponds to, respectively, 40, 76.3, and 21.8%. Taboo slips were produced by 17 participants while neutral slips were produced by 12 participants.

³ As concerns the fake memory tests, results were not exploitable. The results were very variable: some participants remembered word pairs while others only remembered single words which could have appeared in different trials, the number of word (pairs) recalled by the participants was very variable (between 1 and 15) and some participants remembered words that were not part of the stimuli that were presented. These results were not surprising given the large amount of words presented during the SLIP-task. When counting only word couples, there were not enough results, with a lot of missing results in participants, to make statistical analyses. We also did not notice any remarkable qualitative results (e.g., parapraxes emerging or recurring in the memory test).

Note that these 37 parapraxes were associated with 11 sequences out of the 32 in the neutral condition and 7 sequences out of 32 in the taboo condition. As concerns the neutral parapraxes, 2 sequences produced 7 out of 16 parapraxes: *pomme roche* (apple rock, $N = 4$) and *pige fil* (understands wire, $N = 3$), the other 9 parapraxes were produced by 9 sequences and 21 sequences produced no parapraxes. As concerns the taboo trials, 7 sequences out of 32 led to spoonerisms, with 3 “star” sequences: *belle pipe* (nice blowjob, $N = 9$), *bite molle* (limp dick, $N = 4$), and *bite chaude* (warm dick, $N = 4$). The 25 other sequences produced no parapraxes.

3.2. People make as many “taboo” as “neutral” slips

The descriptive statistics (see Table 1) show that there is no significant difference between the occurrence of taboo and neutral slips (Wilcoxon Signed Ranks test; $p = 0.348$), and even not between taboo and neutral speech errors (Wilcoxon signed Ranks test; $p = 0.104$), even though in absolute numbers there are more taboo slip and speech errors. As concerns the partial slips, there was also no difference between the production of partial neutral or taboo spoonerisms (resp. 13 on 16 and 14 taboo on 21; Wilcoxon Signed Ranks test; $p = 0.861$). Among the partial taboo spoonerisms, Motley et al. (1981a, 1982) further distinguished between “safe partials” (e.g., *tool kits* → *cool kits*) and “taboo partials” (e.g., *tool kits* → *tool tits*); there was no difference between the production of safe and taboo partials (resp. 10 and 4, Wilcoxon Signed Ranks test; $p = 0.153$). Interestingly, there were significantly more taboo than neutral omissions (Wilcoxon Signed Ranks test; $p = 0.005$).

Similarly, the category of slips (neutral or taboo) did not explain the occurrence of the slips (Mann–Whitney U -test; $p = 0.488$). We also used the empirical tabooeness data instead of the *a priori* taboo/neutral classification. The production of slips was also not explained by the tabooeness of the intended slips (e.g., *cool tits*; simple linear regression $F_{(1,62)} = 1.596$; $p = 0.211$). As people react upon the eliciting pairs, not necessarily upon the intended taboo puns, we also verified the production of slips in function of the tabooeness of the eliciting pairs (e.g., *tool kits*), but this was also not significant (simple linear regression $F_{(1,62)} = 0.679$; $p = 0.413$). These analyses confirm the descriptives: there is no difference in the occurrence frequency of neutral versus taboo slips; we will thus treat all parapraxes as one group.

3.3. The production of parapraxes is a defensive process

According to our hypothesis, the production of parapraxes is a defensive process, and therefore we expect to predict their occurrence in function of defensiveness parameters, both primary process (number of ATT on the GeoCat) and secondary process (number of N-choices in the PN WordList). The descriptives of these parameters (as well as of the MCSD) are given in Table 2.

When we test this model for all parapraxes, the overall regression was significant: $F_{(2,51)} = 8.841$; $p < 0.001$ with an $R^2 = 25.7\%$. The occurrence of parapraxes is significantly explained

TABLE 1 Mean $\pm$ SEM by participants ($N = 55$) for the total number of slips, speech errors and omissions in both neutral and taboo conditions.

	Total	Neutral	Taboo	p
Slips	0.67 $\pm$ 0.13 (37 - 1%)	0.29 $\pm$ 0.09 (16 - 0.4%)	0.38 $\pm$ 0.08 (21 - 0.6%)	0.348
Errors	2.90 $\pm$ 0.33 (160 - 4.6%)	1.25 $\pm$ 0.17 (69 - 2%)	1.65 $\pm$ 0.22 (91 - 2.6%)	0.104
Omissions	0.68 $\pm$ 0.28 (37 - 1%)	0.13 $\pm$ 0.07 (7 - 0.2%)	0.55 $\pm$ 0.22 (30 - 0.8%)	0.005

Total absolute frequencies are given into parentheses; p -values for bilateral testing.

by the number of ATT ($B = 0.186$; $p = 0.003$) and the number of N/PN ($B = 0.099$; $p = 0.007$).

We also tested the whole model for the omissions and for the verbalization errors, to investigate whether a defensive process is also implied in the omissions and verbalization errors. This was not significant, for omissions ($F = 0.009$; $p = 0.991$) nor for verbalization errors ($F = 0.522$; $p = 0.596$).

3.4. Highly defensive people produce more paraphrases than lowly defensive people

Following Weinberger et al.'s (1979) basis for his "repressor's" taxonomy as well as Davis and Schwartz (1987) and Furnham et al. (2003) we isolated lowly and highly defensive participants. Following authors such as Kraft (1998), Ringel (1999), Erskine et al. (2007), or Lévesque et al. (2010), participants' scores were dichotomized at the median (18) to define participants into lowly defensives (LD; $N = 25$; mean_{MCS}D = 14.0 $\pm$ 0.5) and highly defensives (HD; $N = 30$; mean_{MCS}D = 20.4 $\pm$ 0.4; $p = 0.001$ with LD). If we now predict the occurrence of paraphrases in function of the defensiveness category of the participants, the simple linear regression model is significant: $F_{(1,53)} = 2.884$ ($p = 0.048$, unilateral testing; $B_{MCS}D = 0.427). Indeed, highly defensive participants produced double as many slips as lowly defensive participants (Mann-Whitney U -test = 273; $p = 0.025$, unilateral testing; see Table 3 and Figure 3).$

3.5. Only highly defensive people mobilize inhibitory-type defenses for the production of paraphrases: Running away from ambiguity, they stumble upon their words

Since the defensive dynamics might differ qualitatively in highly and lowly defensive people, in agreement with our hypotheses, we now test our model in lowly and highly defensives separately.

TABLE 2 Mean $\pm$ SEM for the number of ATT-choices in the GeoCat (on 6), the number of N-choices in the PN WordList (hence N/PN; on 20) and the Marlowe-Crowne Social Desirability scale (on 33); $N = 55$.

	Mean $\pm$ SEM	Min	Max
ATT	1.5 $\pm$ 0.3	0	6
N/PN	2.0 $\pm$ 0.4	0	14
MCS	17.5 $\pm$ 0.5	9	25

In Table 3, the descriptives of the defense parameters (ATT and N/PN) are also given for the lowly and highly defensive participants of our study. At this level, there are no significant differences. However, the regression results explaining the factors for the production of paraphrases in function of the defense parameters are different in both populations.

For the lowly defensive participants ($N = 25$), the multiple linear regression model for all paraphrases with the number of ATT and N/PN is non-significant: $F_{(2,21)} = 1.754$; $p = 0.198$. However, when removing the N/PN parameter, the regression with only the ATT-parameter ($B = 0.164$) is significant: $F_{(1,22)} = 3.613$; $p = 0.036$; $R^2 = 14.1\%$ (unilateral testing). For lowly defensive participants, only the primary process predicts the occurrence of paraphrases.

For the highly defensive participants ($N = 30$), the multiple linear regression model for all paraphrases with the number of ATT and N/PN as predictors, shows a significant effect: $F_{(2,27)} = 6.601$; $p = 0.005$; $R^2 = 32.8\%$. Both ATT ($B = 0.210$, $p = 0.019$) and N/PN ($B = 0.112$; $p = 0.011$) predictors are significant. Interestingly, in highly defensive participants both primary and secondary process defenses predict the occurrence of paraphrases.

Note that we also compared lowly (LD; $N = 25$) and highly defensive (HD; $N = 30$) participants for the omissions and the verbalization errors. We found that the ATT and N/PN predictors were not significant for the omissions (LD: $F = 0.307$; $p = 0.739$; HD: $F = 0.083$; $p = 0.920$) nor for the verbalization errors (LD: $F = 0.974$; $p = 0.394$; HD: $F = 0.168$; $p = 0.847$).

3.6. What about differences between neutral and taboo paraphrases?

Even if our data show that the *a priori* difference between neutral and taboo paraphrases is not reflected in empirical differences in treating these paraphrases, for reasons of comparability with previous studies, we have explored this distinction nevertheless. Highly defensive participants produced significantly more taboo spoonerisms ($\mu = 0.53 \pm 0.68$; Mean rank_{HD} = 31.40) than lowly defensive participants ($\mu = 0.20 \pm 0.50$; Mean rank_{LD} = 23.92; Mann-Whitney U -test; $p = 0.033$). However, for the number of neutral spoonerisms no significant differences between highly ($\mu = 0.33 \pm 0.60$; Mean rank_{HD} = 29.30) and lowly defensives ($\mu = 0.24 \pm 0.66$; Mean rank_{LD} = 26.44) were found (Mann-Whitney U -test; $p = 0.360$).

As concerns the regression results explaining the factors for the production of paraphrases in function of the defense parameters in both populations, for neutral paraphrases, we find the same differences as already found for all the paraphrases; interestingly, nothing comes out of the regression analyses for the taboo paraphrases:

TABLE 3 Mean $\pm$ SEM for the number of slips, of verbalization errors and of omissions as well as for the number of ATT-choices in the GeoCat (on 6) and of N-choices in the PN WordList (on 20) by participant, for the lowly and highly defensive participants (LD and HD resp.); p -values for bilateral testing.

	Total population ($N = 55$)	LD ($N = 30$)	HD ($N = 25$)	p
Slips	0.67 ± 0.13 (37)	0.44 ± 0.17 (11)	0.87 ± 0.18 (26)	0.050
Errors	2.90 ± 0.33 (160)	2.32 ± 0.41 (58)	3.40 ± 0.49 (102)	0.133
Omissions	0.68 ± 0.28 (37)	1.08 ± 0.58 (27)	0.33 ± 0.18 (10)	0.292
ATT	1.5 ± 0.3	1.4 ± 0.4	1.6 ± 0.3	0.327
N/PN	2.1 ± 0.4	1.3 ± 0.4	2.8 ± 0.7	0.133

- For the lowly defensive participants ($N = 25$), a simple linear regression for *neutral slips* with ATT as a predictor is significant: $F_{(1,22)} = 5.968$; $p = 0.023$; $R^2 = 21.3\%$ with for ATT: $B = 0.155$. The linear regression in this group for *taboo slips* with ATT is not significant: $F_{(1,22)} = 0.033$; $p = 0.857$ (with for ATT $B = 0.010$).
- For the highly defensive participants ($N = 30$), the multiple linear regression for *neutral slips* with ATT and N/PN as predictors, shows a significant effect: $F_{(2,27)} = 6.606$; $p = 0.005$; $R^2 = 32.9\%$. Both ATT ($B = 0.146$; $p = 0.010$) and N/PN ($B = 0.062$; $p = 0.022$) predictors are significant. The multiple linear regression in this group for *taboo slips* with ATT and N/PN is not significant: $F_{(2,27)} = 1.507$; $p = 0.240$ (with for ATT, $B = 0.064$; $p = 0.343$; and for N/PN, $B = 0.049$; $p = 0.148$).

4. Discussion

Altogether, 37 parapraxes were produced in this study, by 22 participants out of the 55. Although these 37 parapraxes represent about 1% of the experimental trials,⁴ this number is comparable to other studies having replicated the SLIP technique (e.g., Hartsuiker et al., 2005: 6.5%; Costa et al., 2006: 1.9%; Severens et al., 2011: 2.4%; Wagner-Altendorf et al., 2020: 0.7%). The proportion of partial spoonerisms in the total number of spoonerisms is high (more than 70%), but it is comparable to what has been reported in other studies (ca. 47% in Costa et al., 2006 and 68% in Hartsuiker et al., 2006). Interestingly, these frequent partial spoonerisms echo with Freud's claims on verbal parapraxes, namely that they are « an outcome of a compromise: they constitute a half-success and a half-failure for each of the two intentions; the intention which is being challenged is neither completely suppressed nor, apart from special cases, carried through quite unscathed » (Freud, 1916–1917/1966, p. 66).

As concerns the nature of the parapraxes, the 2 “star” parapraxes responsible for more than 40% of the neutral parapraxes—*pomme roche* (apple rock) and *pige fil* (understands

wire) – show a repetition of the middle vowel that might have facilitated these parapraxes (Motley and Baars, 1976a). Similarly, the fact that/b/and/p/are consonants are close in acoustic and articulatory characteristics has probably facilitated the taboo parapraxis *belle pipe* (nice blowjob) responsible in and by itself for more than 40% of the taboo parapraxes. Even if the taboo top 3 sequences are three “penis”-stimuli (next to *belle pipe* there was also *bite molle–limp dick* - and *bite chaude–warm dick*), which might suggest that the “penis”-meaning acts as an amplifying factor, it should be said, though, that the parapraxes *bite frotte* (dick rubs), *bande fort* (big boner), *belle couille* (nice nuts), and *couille molle* (half-sack) were not produced, weakening this proposition. Therefore, we have no strong indication for the nature of the meaning universally determining the probability of making slips. In conclusion, even if this was not the focus of the present research, it seems that elements concerning the phonological nature of the stimulus material influence the probability of making slips, which is in agreement with linguistic research (Motley and Baars, 1976a). Even if our study investigates what in the personality of the participants makes the tongue slip, we do not expect that personality structure covers the whole (or even the major) part of the variability, leaving room for other factors, including the phonological nature of the stimulus material.

4.1. People make as many “taboo” as “neutral” slips

One main finding is that, contrary to Motley et al.'s, recurrent finding that people make less taboo than neutral slips, we did not find such a significant difference between taboo and neutral slips. To take into account that there might be a different understanding (depending on time and place) of what is “taboo,” we measured tabooess in an independent sample of $N = 120$ participants, parallel to the SLIP study. This study confirmed that, on average, the taboo spoonerisms (e.g., *cool tits*) were indeed rated as significantly more taboo than the neutral spoonerisms (e.g., *barn door*). However, tabooess varied considerably, and overlapped partially between *a priori* taboo and *a priori* neutral pairs (the overlap was in the tabooess-range 1.8–2.1). In fact, one taboo spoonerism, *seins mère* (mother's boobs) fell into the range of the “neutral” word pairs (tabooess = 2.11). For this reason, next to using the *a priori* taboo/neutral categories, we also regressed the probability of making a parapraxis upon the tabooess index, but this regression was also not significant. Finally, we did the same exercise with the tabooess of the eliciting pairs, with no significant

⁴ We discussed this point with Michael Motley and in a personal communication (April 2020), he explained that « we had trouble ourselves when trying with presentation *via* computer. The memory drum was slightly noisy and had an audible rhythm that created a metronome effect. I don't know for certain, but I think that the “ka-chunk” sound was an important part of the SLIP technique ». This might be true. However, even if the memory drum might have enabled more slips, the laboratory conditions are close to the slips occurring naturally and do not prevent from doing statistical analyses.

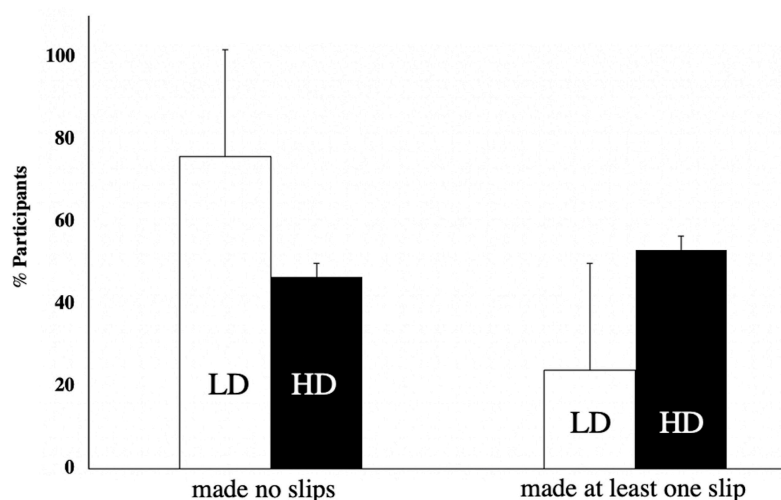


FIGURE 3

Main behavioral effects: percentage of number of lowly ($N_{\text{total}} = 25$) and highly ($N_{\text{total}} = 30$) defensive participants who made no slips versus at least one slip ($\chi^2_{(1)} = 4.889$; $p = 0.027$; the χ^2 test indicates that the total correlation between “making no or at least one slip” and “being lowly or highly defensive” is significant).

results. By all means, then, we may say that there were no significant differences between the numbers of taboo and neutral paraphrases.

Severens et al. (2011) also did not find significant results. Both in our study and in Severens et al. (2011) absolute numbers were even *higher* for taboo slips than for neutral slips, but in both studies the difference was not significant. To our knowledge, only Wagner-Altendorf et al. (2020) have replicated Motley et al. (1981a, 1982) original results, finding significantly more neutral than taboo slips. Motley et al. (1982, p. 580) had interpreted their findings of fewer taboo slips as an indication for the existence of a prearticulatory editing component that prevents the overt formulation of taboo words on the basis of « social appropriateness ». Having not replicated their results, we do not, however, interpret these findings as a disproof of Motley’s proposed “editor” or “censor”-principles. Indeed, we had initially predicted, following herein Freud’s repeated injunctions (see Introduction), that all slips, even the « apparently simple » (Freud, 1901/1978, p. 83) or « the mildest cases » (Freud, 1901/1978, p. 279), were to be understood as the result of a defensive process, and that the *a priori* difference between “neutral” and taboo is not valid at a singular subjective level.

4.2. The production of paraphrases is a defensive process

Moreover, our results are in line with Freud’s hypothesis that in general paraphrases are to be understood as a result of defensive mechanisms (see Introduction: Freud, 1901/1978, p. 279). With our two parameters, the ATT in the GeoCat, which is thought to “catch” primary process mentation, and the N/PN parameter, which is supposed to “catch” a secondary process defensive move against linguistic ambiguity, our model captures a big fourth of the variability in the production of paraphrases in a significant way (with both parameters being significant). In other words, both an increase in primary process mentation

and an increase in defensive avoidance of language ambiguity significantly predict the occurrence of paraphrases, confirming our first hypothesis. Furthermore, these ATT and N/PN parameters did not significantly predict the number of omissions and verbalization errors. Therefore, only paraphrases lend themselves to analysis and interpretation in terms of mental categories confirming the Freudian position that slips of the tongue are not simply “system glitches” but subjectively intentional mental phenomena.

We are, of course, not overly amazed to catch only a quarter of the variance of the paraphrases production. Our expectation is that the by chance-correspondence of the presented meanings with the singularly important meanings of the specific subject will actually catch the big chunk of the variance. Our bet was that, beyond the singularity of the meanings, universal formal logics do play a role in explaining or predicting the probability of producing paraphrases. Indeed, it appears that not everything concerning our mental productions is a question of the meanings which inhabit our subjective life. There are formal organization logics which structure this world of meanings and, moreover, which may structure them differently according to personality. For example, Bergeret, (1974, p. 46) has described a personality in terms of a « rather invariant reciprocal play of the primary and secondary processes ». For all these reasons, following Freud, clinical experience, and in accordance with others, defensiveness and its articulation as a (differential) combination of primary and secondary process mentation are used here as the key principles to map these different organizations.

4.3. Highly defensive people produce more paraphrases than lowly defensive people

We distinguished two populations, lowly and highly defensive participants, in the general population based upon their results

on the Marlowe Crowne Social Desirability scale. We have argued that although this scale is originally designed to measure social desirability, it has almost directly and consistently since been used by the original authors (Crowne and Marlowe, 1964) and by others (e.g., Weinberger et al., 1979; Weinberger, 1990; Eysenck and Van Berkum, 1992; Mann and James, 1998), as a measure of defensiveness. Still, as such it has remained a controversial measurement tool because of its ambiguous factorial structure (e.g., Leite and Beretvas, 2005). For this reason, we have measured, in a separate population, its convergent and divergent validity with another validated tool for social desirability, as well as with a validated “repressive defensiveness” scale and with a validated anxiety inventory, all yielding coherent results (Bruchon-Schweitzer and Paulhan, 1990; Tournois et al., 2000; Paget et al., 2010). Nevertheless, our Cronbach’s Alpha in the present study was not good, and this is one of the limitations of this study. The low N in the present study (55) probably explains this low Cronbach’s Alpha, as our parallel variable evaluation study for the MCSD with a much larger population yielded a good Cronbach’s Alpha. In this respect, the eigenvalue of the first factor in the Principal Component Analysis (PCA) of the SLIP-MCSD was 3.6 and Yurdugül (2008) underscore that with only one factor with an eigenvalue between 3 and 6, a sample of *at least 100* is needed to reliably calculate the Cronbach’s Alpha. Furthermore, in the present study, we only used the SLIP-MCSD results to divide our population in lowly and highly defensive participants and our criterion (the median of 18) is comparable in absolute numbers to the criterion used in other studies: e.g., Weinberger et al. (1979), Fuller and Conner (1990), and Weinberger and Davidson (1994) divided their population at resp. 17, 18, and 19, using the upper quartile as a criterion. Finally, even with this weak Cronbach’s Alpha, the dichotomic categorization on the MCSD significantly predicted the occurrence of parapraxes in function of the defensiveness category, empirically confirming the relevance of this distinction in the population. For these reasons, we think that our categories of lowly and highly defensive participants are valid categories. Our results now show that highly defensive parameters make almost double as many parapraxes compared to lowly defensive parameters (see Figure 1), but do *not* make more speech errors or more omissions. This confirms the link between defenses and parapraxes. To further explain this result, we have investigated the primary and secondary process logics linked to the occurrence of parapraxes in each population.

4.4. Only highly defensive people mobilize inhibitory-type defenses for the production of parapraxes: Running away from ambiguity, they stumble upon their words

Our results show that the occurrence of parapraxes in lowly defensive parameters is only significantly explained by the ATT primary process parameter and with a low R^2 (14%). The N/PN parameter does not contribute to explaining the variance in the occurrence of parapraxes. In highly defensives, however, the model with both parameters explains almost a third of the

variance, with both the primary and secondary process parameter being significant.

The low contribution in explained variance in the lowly defensive parameters does not mean, in our opinion, that their defenses are necessarily low. As said in the Introduction, we consider, with Freud, that the ontological nature of mental processes is defensive in essence. In the primary process logic, the defense consists in directly associatively discharging the word (pairs) that have gathered a high tension - this is, without this tension being first built up to stocked excitation by means of inhibition. To defend against the excitement caused by the possibility of saying *cool tits* when reading *tool kits*, one simply directly says *cool tits*. This direct and transparent way of dealing with uptight topics is in colloquial language, paradoxically, sometimes qualified as “non-defensive,” especially in its opposition with inhibitory, restrictive defenses. This should not obscure the basic fact, however, that it also is defensive. The obsessional preoccupation with or projection of the same topic over and again, it being thereby positively present as Erdelyi (2006) would say, shows in its exaggeration that primary process directness is, in essence, also defensive. However, we think we could not catch this primary process logic in an effective way due to the fact that we might not have an adapted measurement tool for primary process linguistic mechanisms.

In the highly defensive participants, in contrast, we succeeded in capturing a good part of the explained variance with our two parameters. However, it is important to point out that the distinction between lowly and highly defensives was made on the basis of the extent to which participants were inclined *not to acknowledge* undesirable social behavior. In other words, highly defensive participants are thought to specifically have high defenses of the second category, the inhibitory category - this is, the category linked by us to the secondary process. This means that our findings might at first sight look somewhat circular: we selected participants with high inhibitory defenses and then find that indeed their parapraxes are explained by these inhibitory defenses. However, we must remember that our measure for “these inhibitory defenses” is quite radically different from the MCSD-personality questionnaire, as it concerns a very basic linguistic similarity preference (between a phonologically similar and a non-similar choice). It is therefore actually rather revealing that participants deemed defensive on the basis of the MCSD are also making more parapraxes in correlation with the avoidance of language ambiguity. In other words still, it is remarkable to find here empirical evidence for the idea that psychodynamic defenses show up as a specific manner of processing language (see also Lacan, 1966), a manner which we may catch by the phrase: with increasing defensiveness, we run away from ambiguity and stumble upon our words. Indeed, the avoidance of subliminal ambiguity was found before (see Bazan et al., 2019a); moreover, here this avoidance is linked to a higher probability to make parapraxes, i.e., to stumble upon words.

In summary, defensive dynamics are both similar and different in lowly and highly defensive parameters. They are similar in that primary process defenses might possibly be seen as baseline default defenses, but different in that highly defensive participants mobilize an additional mechanism, which, as we propose, is inhibitory in nature and is secondary process in type.

4.5. What about differences between neutral and taboo parapraxes?

Even if the difference between neutral and taboo parapraxes is an artificial, non-empirically confirmed distinction, when we apply these *a priori* categories, we find that highly defensive participants produce more than double the number of taboo parapraxes produced by lowly defensive participants, while they did not produce significantly more neutral parapraxes, even if the number were higher in absolute terms. Interestingly, understanding the production of taboo parapraxes selectively in terms of primary and secondary processes does not work: in lowly defensive parameters, where the model only takes the primary process, results are nowhere and in highly defensive participants, where the model takes both primary and secondary process parameters, the results are not significant. In contrast, the production of neutral parapraxes in lowly and highly defensive participants gives results in line with those for the total group of parapraxes: in lowly defensives the model is significant with the AIT accounting for a good 20% of the explained variance; in highly defensives the significant model accounts for a third of the variance, with both AIT and N/PN being significant.

This, then, brings us to the following speculation: indeed, the universally taboo parapraxes form a distinct group, and indeed, what distinguishes them from the universal category of “neutral” parapraxes, is that they are as a group more subject to an editing process. However this editing process pointed out by Motley et al. (1981a, 1982) is, in our view, primary nor secondary process-type. This is, it is not an *unconscious* defense type; it is not repression, but a (pre-)conscious *suppression* mechanism. Some other results point in that direction. Indeed, impressively, in our parallel study on the tabooeness of the word pairs implied in the SLIP-task, - namely, the spoonerism eliciting pairs and the spoonerisms themselves - subjects rated taboo eliciting pairs (e.g., *tool kits*) as being more taboo than neutral eliciting pairs (e.g., *darn bore*). We propose that these results show that people are capable of anticipating taboo slips. This might suggest that there is a preconscious intuition of the taboo outcome of words that are nevertheless neutral at first sight. The difference between suppression and repression is made explicit in a superb way by Freud (1901/1978) in his explanation of the forgetting of the name “Signorelli.” The first substitute, which comes to his mind, instead of Signorelli, is *Botticelli*. He explains this substitution by suggesting that the mental cathexis of “signor,” under inhibition, has by the way of its translation to “Herr,” followed by the word-bridge “Herzegowina und Bosnien,” migrated to the syllable “Bo,” which, together with the uninhibited syllable “elli,” leads to Botticelli. Crucially, he indicates that he had consciously swallowed a phrase he was about to say to his travel companion and which started as “Herr, was ist da zu sagen” (“Sir, what is there to be said?”). He remembered having done so since it referred to inappropriate sexual content. He did so well in withholding this, that the inhibition spilled over to the associated fragments and he also inhibited the associated semantic and phonological variants, including the Italian translation of *Herr*, *signor*, and the phonologically associated *Signorelli*. But since the move is consciously remembered, this is suppression, not repression. However, another substitute comes to his mind, *Boltraffio*. Associating upon this substitute, Freud, (1901/1978, p. 3)

now remembers that one of his patients over whom he « had taken a great deal of trouble », had committed suicide and that he came to know this information when he was in Trafoi. It makes sense that the chain of associations, drifting upon the general theme “death and sex,” had also activated *Trafoi*, which, inheriting the highly cathected “Bo,” could find discharge, disguised as “Boltraffio.” However, Freud had no conscious recollection whatsoever that this patient had come to his mind and only reconstructed this probability in the aftermath on the basis of the substitute word. This, then, indicates that the “Trafoi”-associations were properly repressed and Boltraffio is the return-of-the-repressed.

What we propose, then, is that the Motley prearticulatory editor indeed intercepts taboo parapraxes in such a way that there are significantly more taboo-parapraxes that were swallowed in a final editing process than neutral parapraxes. We propose that this “swallowing” is a late censorship, proximal to utterance and independent of the intimate personality organization, and therefore not explainable in terms of primary process and secondary process inhibition, as seen in our results. Two independent results give more weight to this hypothesis. First, the number of omissions in taboo target pairs were more than five times the number of omissions in neutral target pairs (see Table 1), suggesting people might more often “swallow” a taboo parapraxis they are about to make. Also, the number of omissions was not explainable in terms of primary and secondary processes mentation as the model was not significant.⁵ Second, there was no significant difference in the number of omissions between lowly and highly defensive participants (in absolute numbers, the highly defensives had even less), indicating that indeed the “swallowing” was not an unconscious defense move, explainable in terms of primary process and secondary process inhibition, but indeed an editing process of another nature, comparable in psychoanalytic terms to conscious or preconscious suppression. It might also be called a “cognitive correction” which happens quite independently of personality (Shevrin, 1992).

This interpretation would also fit nicely with the other SLIP-studies, including a number of more recent neuro-imaging results. First, let’s recall Motley et al. (1981a, 1982); p. 9, p. 580 editor is described as « prearticulatory editing on the basis of social appropriateness », much along what Freud, (1901/1978, p. 3) himself recount as concerns his swallowing of the « Herr, was ist da zu sagen »-anecdote: « I suppressed my account of this characteristic trait, since I did not want to allude to the topic in a conversation with a stranger » (Freud, 1901/1978, p. 3). As said, we propose that the censorship which is at play here is not the censorship between the system Ucs and conscious processing (namely, repression) but rather, the second censorship, the one between the Pcs and the Cs: « a mental act commonly goes through two phases, between which is interposed a kind of testing process (censorship). In the first phase the mental act is unconscious and belongs to the system Ucs; if upon the scrutiny of the censorship it is rejected, it is not allowed to pass into the second phase; it

⁵ A total of 12 out of 55 participants actually made omissions. Anecdotally, 13 of the 37 total omissions were made by only one participant. These omissions appeared shortly after the participant made a taboo parapraxis - *belle pipe* (nice blowjob) - and 10 of the 13 omissions indeed concerned taboo eliciting trials. This confirms the idea of conscious swallowing, here probably after being alarmed by the own first (taboo) parapraxis.

is then said to be “repressed” and must remain unconscious. If, however, it passes this scrutiny, it enters upon the second phase and thenceforth belongs to the second system, which we will call the Cs. But the fact that it is so belongs does not unequivocally determine its relations to consciousness. It is not yet conscious, but it is certainly *capable of entering consciousness*, (...) that is, it can now, *without any special resistance* and given certain conditions, become the object of consciousness. In consideration of this capacity to become conscious we also call the system Cs the “*preconscious*” » (Freud, 1915/1963, pp. 122–123; Italics added). We propose that a number of slips, having passed the censorship of repression, and being readied to be discharged, i.e., executed or articulated, are halted at that stage: to cite Freud, (1915/1963, p. 123), even if they were « capable of entering consciousness », they encountered « special resistance », and were censored from the system Cs; this particular resistance would then precisely be Motley et al.’s socially motivated prearticulatory editor.

Severens et al. (2011) found that, slightly after being confronted with the speech prompt, participants showed a larger negative brain wave in the taboo condition compared to the neutral condition, even if they produced no spoonerisms (see also Wagner-Altendorf et al., 2020). The authors interpreted this as evidence that taboo errors « are formed, detected, and corrected internally » and as « the first direct evidence that covert editing of speech exists » (Severens et al., 2011, pp. 1256–1257) which might also constitute evidence that *preconscious* editing in speech exists. In 2012, the same authors show that the inhibition of taboo words activates the right inferior frontal gyrus (rIFG) – a region which might implement neural inhibition of manual (Aron et al., 2003, 2007; Chambers et al., 2006) and verbal responses (Xue et al., 2008). Furthermore, the rIFG has been previously associated with *externally* triggered inhibition (see Severens et al., 2012; e.g., control of risky behavior and delayed gratification as well as emotion regulation). Severens et al., (2012, p. 431) comment: « This finding strongly suggests that external social rules become internalized and act as a stop-signal » and refer to the fact that it is through education and socialization, i.e., through external signals, that we have learned to inhibit inappropriate behavior, as well as not to utter taboo words.

On the other hand, inhibition that is not guided by an external cue but rather internally guided (endogenous self-control), has been demonstrated to involve the dorsal fronto-median cortex (dFMC; Brass and Haggard, 2007, 2008; Kühn et al., 2009). As pointed out by Severens et al., (2012, p. 431), it might be « very crucial to distinguish endogenous from externally guided inhibition » conceptually, confirming also their neuroanatomical distinction. However, we disagree that « external guided inhibition is not a result of deliberation but is rather triggered by the environment » while « by contrast, endogenous self-control is related to a deliberate decision » (Severens et al., 2012, p. 431). We think that *any instance* of inhibition is a result of (subjective) deliberation. When it comes to « externally inhibited » behavior, this deliberation is relatively easy to access consciously, and might correspond to the censorship between the systems Pcs and Cs in the psychoanalytic model. Endogenous self-control, on the other hand, concerns control over internally motivated action, of which a large part is conscious but of which, we propose, another (considerable) part remains unconscious. Freudian repression is supposed to have this characteristic that it happens without any conscious awareness whatsoever and that is very difficult to become aware of, and

therefore also to record in experimental set-ups. As seen, it often involves logical suppositions and *post hoc* reconstructions similar to the “Boltraffio/Trafoi”-reconstruction. The self-control decision not to act upon certain endogenous urges, even if deliberate, might thus happen completely unconsciously (which then would constitute an instance of repression) while still activating (dFMC) brain areas involved in decision making (Brass and Haggard, 2007, 2008; Kühn et al., 2009).

These different considerations suggest that the Motley prearticulatory editor remains an *external* instance of inhibition. This external nature pertains to the actual role of others, of the social realm, of the ones we relate to in the very moment – see also the different ways this “actual other” was operationalized in different SLIP-studies [e.g., a sexually provocative experimenter, in Motley and Baars (1978)]. In our view, endogenous self-control pertains to the intimate realm of singularly specific meanings, out of which the endogenously motivated actions spring. As the prearticulatory editor guarantees the social norms, even when internalized, it remains the internal representative of these social norms; it does not relate to the intimate realm of singularly specific meanings of the subject. Concretely, independently of one’s life story *belle pipe* (*nice blowjob*) is taboo, but if you had a nasty experience in a *lighthouse* then *mauve phare* (*purple lighthouse*) might be to you specifically very taboo. In our results, we saw how the slips as a total group, and specifically the so-called “neutral slips,” were to a certain extent predictable in terms of the personality-specific mix of primary process and secondary process inhibition. To the contrary, the taboo spoonerisms did not form anything like a coherent group in relation to these mental categories – in other words, they did not as a group acquire a mentally intimate taboo significance (independently of the personality structure of the participants), and for this reason were not as a group more subject to repression than other slips.

However, given the higher number of taboo parapraxes in highly defensive participants (ca. 2.65 times the number in lowly defensives), we must assume that even if a number of taboo parapraxes were “luckily” intercepted in time – this is, before utterance – still a number escaped vigilance, and that this number was higher in the highly defensives. It would be logical to think that conflictual themes are under higher pressure in highly defensives, so that when the opportunity to express these themes arises – as in the SLIP task – their tongue tends to slip more, giving them an opportunity to release some mental pressure and to avoid tension accumulation. This, then, is precisely Freud’s (1915/1963) « return-of-the-repressed ». It also means that, next to being more regulated by socially internalized prohibitory rules, taboo stimuli also must have a higher probability to be conflictual, even if only so for the highly defensive participants.

Our study shows experimental evidence to support a psychodynamic explanatory model for the production of parapraxes and more widely that psychodynamic (Freudian) phenomena lend themselves in a refutable way to experimental research. Together with previous experimental studies on subliminal language ambiguity (Bazan et al., 2019a) and on the unwitting resolution of rebuses (Olyff and Bazan, 2023), imbedded in a clinically inspired theoretical model (e.g., Bazan, 2007, Bazan, 2012; Bazan and Snodgrass, 2012; Bazan et al., 2021), this study contributes to the critical development of the psychoanalytic corpus and its fruitful integration into the scientific

corpus of adjacent disciplines, such as cognitive neuroscience, psycho- and neurolinguistics.

5. Limitations

One of the limitations we already mentioned and discussed (see higher) is the low value for the Cronbach's Alpha of the MCSD in the present study. Another limitation is that the parallel study on the subjective evaluations for the tabooess of the word pairs presented in the SLIP-task were done by external participants.

6. Conclusion

In conclusion, our research in 55 French speaking participants with 32 taboo and 32 neutral paraphrases, administered through the SLIP method (e.g., Motley and Baars, 1976a) shows that, contrary to previous results, people do not make more taboo than neutral paraphrases and that, in line with Freud's ideas, all paraphrases can be partially explained in terms of defenses, both of the elaborative primary process and of the inhibitory secondary process kind (Erdelyi, 2006). Splitting up the population in lowly and highly defensives proved productive as it shows that only in highly defensive people the production of paraphrases also involved a secondary process type of defense against language ambiguity. In other words, the more we run away from language ambiguity, the more we stumble upon words. This kind of findings contribute to psychoanalytic knowledge by enabling to experimentally back-up a Freudian model of repression and of return of the repressed. At the same time, our results corroborate the existence of the prearticulatory editor for the taboo words, but situate it at an external locus of control, independent of the intimate singular mental life, and comparable to the censorship between the systems Preconscious and Conscious in a metapsychological model.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the "Comité d'Avis Éthique de la Faculté des

Sciences Psychologiques et de l'Éducation" of the Université Libre de Bruxelles. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AB had the idea for the research. LT and LP created the French SLIP-task stimuli, operationalized the protocol on PsychoPy, conducted the experiments, and carried out the data analysis. AB, SD, and GO provided supervision at all stages of the research. LT and AB wrote the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are heartily thankful to Mr. Robert Berry and the International Psychoanalytical Association (IPA) for financial support. AB is thankful for the decisive mentoring received from Howard Shevrin in Ann Arbor (2003–2005) and beyond. We are also thankful for Laurine Sanchez and Yadamjav Chunag Balson for their contribution in the parallel study on the subjective evaluations on the "tabooess" of the word pairs presented in the SLIP-task and to Camille Demarque for her contribution in the parallel study on the MCSD validation.

Conflict of interest

The authors declare that this study received modest funding from the International Psychoanalytical Association (IPA) and from Mr. Robert Berry, enabling functioning expenses. However, neither the IPA nor Mr. Robert Berry were involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., and Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J. Neurosci.* 27, 3743–3752. doi: 10.1523/JNEUROSCI.0519-07.2007
- Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., and Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* 6, 115–116. doi: 10.1038/nn1003
- Bazan, A. (2007). *Des Fantômes Dans La Voix. Une Hypothèse Neuropsychanalytique Sur la Structure De L'inconscient*. Montréal: Liber.
- Bazan, A. (2012). From sensorimotor inhibition to Freudian repression: insights from psychosis applied to neurosis. *Front. Psychol.* 3:452. doi: 10.3389/fpsyg.2012.00452
- Bazan, A., and Brakel, L. A. (2023). The GeoCat 1.3, a simple tool for the measurement of Freudian primary and secondary process thinking. *Neuropsychanalysis* 2, 1–11. doi: 10.1080/15294145.2023.2169955

- Bazan, A., Kushwaha, R., Winer, E. S., Snodgrass, J. M., Brakel, L. A. W., and Shevrin, H. (2019a). Phonological Ambiguity Detection Outside of Consciousness and Its Defensive Avoidance. *Front. Hum. Neurosci.* 13:77. doi: 10.3389/fnhum.2019.00077
- Bazan, A., Jorge Mendes, S., Antoine-Moussiaux, J., and Prys, D. (2019b). Empirical evidence for psychic transparency in pregnancy. *Psychoanal. Psychol.* 36, 239–248. doi: 10.1037/pap0000212
- Bazan, A., and Snodgrass, M. (2012). “On unconscious inhibition: instantiating repression in the brain,” in *Trends in Psychodynamic Neuroscience*, eds A. Fotopoulou, D. Pfaff, and M. A. Conway (Oxford: Oxford University Press), 307–337. doi: 10.1093/med/9780199600526.003.0018
- Bazan, A., Van de Vijver, G., and Caine, D. (2021). “On the role of language motor dynamics for language processing and for mental constitution,” in *Clinical Studies in Neuropsychanalysis Revisited*, eds C. Salas, O. Turnbull, and M. Solms (London: Routledge), 74–104. doi: 10.4324/9781003188551-7
- Bazan, A., Van Draeghe, K., De Kock, L., Brakel, L. A., Geerardyn, F., and Shevrin, H. (2013). Empirical evidence for Freud's theory of primary process mentation in acute psychosis. *Psychoanal. Psychol.* 30, 57–74. doi: 10.1037/a0027139
- Bergeret, J. (1974). *La Personnalité Normale et Pathologique*. Paris: Dunod.
- Brakel, L. A., Kleinsorge, S., Snodgrass, M., and Shevrin, H. (2000). The primary process and the unconscious: experimental evidence supporting two psychoanalytic presuppositions. *Int. J. Psychoanal.* 81, 553–569. doi: 10.1516/0020757001599951
- Brass, M., and Haggard, P. (2007). To do or not to do: the neural signature of self-control. *J. Neurosci.* 27, 9141–9145. doi: 10.1523/JNEUROSCI.0924-07.2007
- Brass, M., and Haggard, P. (2008). The what, when, whether model of intentional action. *Neuroscientist* 14, 319–325. doi: 10.1177/1073858408317417
- Bruchon-Schweitzer, M., and Paulhan, J. (1990). *Inventaire D'anxiété État-Trait Forme Y (STAI-Y)*. Paris: Éditions du Centre de Psychologie Appliquée.
- Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., et al. (2006). Executive “brake failure” following deactivation of human frontal lobe. *J. Cogn. Neurosci.* 18, 444–455. doi: 10.1162/089992906775990606
- Costa, A., Roelstraete, B., and Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: evidence for feedback between lexical and sublexical levels across languages. *Psychon. Bull. Rev.* 13, 972–977. doi: 10.3758/BF03213911
- Crowne, D., and Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *J. Consult. Psychol.* 24, 349–354. doi: 10.1037/h0047358
- Crowne, D. P., and Marlowe, D. (1964). *The Approval Motive: Studies in Evaluative Dependence*. New York, NY: Wiley.
- Davis, P. J., and Schwartz, G. E. (1987). Repression and the inaccessibility of affective memories. *J. Pers. Soc. Psychol.* 52, 155–162. doi: 10.1037/0022-3514.52.1.155
- Dell, G. S., and Reich, P. A. (1980). “Slips of the tongue: the facts and a stratificational model,” in *Papers in Cognitive-Stratificational Linguistics*, eds J. Copeland and P. Davis (Houston, TX: Rice University), 19–34.
- Dell, G. S., and Reich, P. A. (1981). Stages in sentence production: an analysis of speech error data. *J. Verb. Learn. Verb. Behav.* 20, 611–629. doi: 10.1016/S0022-5371(81)90202-4
- d'Épinay, M. L. (2003). Préconscient et fonction du langage. *Libres Cahiers Pour La Psychanalyse* 1, 79–90. doi: 10.3917/lcpp.007.0079
- Dhooge, E., and Hartsuiker, R. J. (2011). How do speakers resist distraction?: evidence from a taboo picture-word interference task. *Psychol. Sci.* 22, 855–859. doi: 10.1177/0956797611410984
- Erdelyi, M. H. (2006). The unified theory of repression. *Behav. Brain Sci.* 29, 499–511. doi: 10.1017/S0140525X06009113
- Erskine, J. A., Kvavilashvili, L., Conway, M. A., and Myers, L. (2007). The effects of age on psychopathology, well-being and repressive coping. *Aging Mental Health* 11, 394–404. doi: 10.1080/13607860600963737
- Eysenck, M. W., and Van Berkum, J. (1992). Trait anxiety, defensiveness, and the structure of worry. *Pers. Individ. Diff.* 13, 1285–1290. doi: 10.1016/0191-8869(92)90170-T
- Freud, S. (1895/1966). “Project for a scientific psychology,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. I, ed. J. Strachey (London: Hogarth Press), 281–392.
- Freud, S. (1900/1955). *The Interpretation of Dreams*. New York, NY: Basic Books.
- Freud, S. (1901/1978). “The psychopathology of everyday life,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. VI, ed. J. Strachey (London: Hogarth Press), 1–279.
- Freud, S. (1914/1964). “On the history of the psycho-analytic movement,” in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. XIV, ed. J. Strachey (London: Hogarth Press), 7–66.
- Freud, S. (1915/1963). “The unconscious,” in *General Psychological Theory, Papers on Metapsychology*, ed. P. Rieff (New York, NY: Macmillan Publishing Company), 115–150.
- Freud, S. (1916–1917/1966). “Introductory lectures on psycho-analysis,” in *The Complete Introductory Lectures on Psychoanalysis*, ed. J. Strachey (New York, NY: W.W. Norton & Company), 1–463.
- Freud, S. (1933/1966). “New introductory lectures on psycho-analysis,” in *The Complete Introductory Lectures on Psychoanalysis*, ed. J. Strachey (New York, NY: W.W. Norton & Company), 465–646.
- Fuller, B. F., and Conner, D. A. (1990). Selection of vigilant and avoidant coping strategies among repressors, highly anxious and truly low anxious subjects. *Psychol. Rep.* 66, 103–110. doi: 10.2466/pr0.1990.66.1.103
- Furnham, A., Petrides, K. V., Sisterson, G., and Baluch, B. (2003). Repressive coping style and positive self-presentation. *Br. J. Health Psychol.* 8, 223–249. doi: 10.1348/135910703321649187
- Green, A. (1995). *Propédeutique: La métapsychologie revisitée*. Seysell: Éditions Champ Vallon.
- Hansen, S. J., McMahon, K. L., and de Zubicaray, G. I. (2019). The neurobiology of taboo language processing: fMRI evidence during spoken word production. *Soc. Cogn. Affect. Neurosci.* 14, 271–279. doi: 10.1093/scan/nsz009
- Hartsuiker, R. J., Antón-Méndez, L., Roelstraete, B., and Costa, A. (2006). Spoonish spanerisms: a lexical bias effect in Spanish. *J. Exp. Psychol.* 32, 949–953. doi: 10.1037/1358-7393.32.4.949
- Hartsuiker, R. J., Corley, M., and Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *J. Mem. Lang.* 52, 58–70. doi: 10.1016/j.jml.2004.07.006
- Hartsuiker, R. J., and Kolk, H. H. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cogn. Psychol.* 42, 113–157. doi: 10.1006/cogp.2000.0744
- Haskell, R. (1991). An analogic methodology for the analysis and validation of anomalous cognitive and linguistic operations in small group (fantasy theme) reports. *Small Group Res.* 22, 443–474. doi: 10.1177/1046496491224003
- Haskell, R. (2001). *Deep Listening. Hidden Meanings in Enveryday Conversation*. Cambridge, MA: Perseus Publishing.
- Haskell, R. E. (2003). A logico-mathematic, structural methodology: part i, the analysis and validation of sub-literal (sublit) language and cognition. *J. Mind Behav.* 24, 347–400.
- Holt, R. (1967). “The development of the primary process: a structural view,” in *Motives and Thought: Essays in Honor of David Rapoport*, ed. R. Holt (New York, NY: International Universities Press), 345–383.
- Jay, T. (1999). *Why We Curse: A Neuro-Psycho-Social Theory of Speech*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/z.91
- Jay, T. (2009). The utility and ubiquity of taboo words. *Perspect. Psychol. Sci.* 4, 153–161. doi: 10.1111/j.1745-6924.2009.01115.x
- Jay, T., Caldwell-Harris, C., and King, K. (2008). Recalling taboo and nontaboo words. *Am. J. Psychol.* 121, 83–103. doi: 10.2307/20445445
- Klein Villa, K., Shevrin, H., Snodgrass, M., Bazan, A., and Brakel, L. A. W. (2006). Testing Freud's hypothesis that word forms and word meaning are functionally distinct: subliminal primary-process cognition and its link to personality. *Neuropsychologia* 44, 117–138. doi: 10.1016/j.neuropsychologia.2006.10.073
- Kraft, M. E. (1998). *The Relationship Between the Repressive Coping Style and Breast Cancer Incidence*. [Doctoral dissertation]. Los Angeles, CA: University of California.
- Kühn, S., Haggard, P., and Brass, M. (2009). Intentional inhibition: how the “veto-area” exerts control. *Hum. Brain Mapp.* 30, 2834–2843. doi: 10.1002/hbm.20711
- Lacan, J. (1966). *Les Écrits I*. Paris: Éditions du Seuil.
- Leite, W. L., and Beretvas, S. N. (2005). Validation of scores on the marlowe-crowne social desirability scale and the balanced inventory of desirable responding. *Educ. Psychol. Meas.* 65, 140–154. doi: 10.1177/0013164404267285
- Levelt, W. J. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/s0140525x99001776
- Lévesque, K., Moskowitz, D. S., Tardif, J. C., Dupuis, G., and D'antonio, B. (2010). Physiological stress responses in defensive individuals: Age and sex matter. *Psychophysiology* 47, 332–341. doi: 10.1111/j.1469-8986.2009.00943.x
- Mann, S. J., and James, G. D. (1998). Defensiveness and essential hypertension. *J. Psychos. Res.* 45, 139–148. doi: 10.1016/S0022-3999(97)00293-6
- Marin, D. (2011). Le symptôme, métaphore ou lettre. *Champ Lacanien* 2, 163–170. doi: 10.3917/chla.010.0163
- Marmursztejn, R. (2013). “Une phobie de signifiant,” in *Les Phobies » Chez L'enfant: Impasse Ou Passage?*, eds M. Bergès-Bouines, S. Calmettes-Jean, C. Ferron, J. M. Forget, and C. Rey (Toulouse: Érès), 107–117.
- Medin, D. L., Goldstone, R. L., and Gentner, D. (1990). Similarity involving attributes and relations: judgments of similarity and difference are not inverses. *Psychol. Sci.* 1, 64–69. doi: 10.1111/j.1467-9280.1990.tb00069.x
- Meringer, R., and Mayer, K. (1895). *Versprechen und Verlesen: eine Psychologische-Linguistische Studie*. Amsterdam: John Benjamins Publishing Company. doi: 10.1515/9783112694480
- Möller, J., Jansma, B. M., Rodriguez-Fornells, A., and Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebr. Cortex* 17, 1173–1178. doi: 10.1093/cercor/bhl028

- Motley, M. T. (1985). Slips of the tongue. *Sci. Am.* 253, 116–127. doi: 10.1038/scientificamerican0985-116
- Motley, M. T., and Baars, B. J. (1976a). Laboratory induction of verbal slips: a new method for psycholinguistic research. *Commun. Q.* 24, 28–34. doi: 10.1080/01463377609369216
- Motley, M. T., and Baars, B. J. (1976b). Semantic bias effects on the outcomes of verbal slips. *Cognition* 4, 177–187. doi: 10.1016/0010-0277(76)90003-2
- Motley, M. T., and Baars, B. J. (1978). Laboratory verification of “Freudian” slips of the tongue as evidence of prearticulatory semantic editing. *Ann. Int. Commun. Assoc.* 2, 141–152. doi: 10.1080/23808985.1978.11923721
- Motley, M. T., Camden, C. T., and Baars, B. J. (1979). Personality and situational influences upon verbal slips: a laboratory test of Freudian and prearticulatory editing hypotheses. *Hum. Commun. Res.* 5, 195–202. doi: 10.1111/j.1468-2958.1979.tb00633.x
- Motley, M. T., Camden, C. T., and Baars, B. J. (1981a). Toward verifying the assumptions of laboratory-induced slips of the tongue: the output-error and editing issues. *Hum. Commun. Res.* 8, 3–15. doi: 10.1111/j.1468-2958.1981.tb00652.x
- Motley, M. T., Baars, B. J., and Camden, C. T. (1981b). Syntactic criteria in prearticulatory editing: evidence from laboratory-induced slips of the tongue. *J. Psycholinguist. Res.* 10, 503–522. doi: 10.1007/BF01076736
- Motley, M. T., Camden, C. T., and Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue. *J. Verb. Learn. Verb. Behav.* 21, 578–594. doi: 10.1016/S0022-5371(82)90791-5
- Murphy, G. L., and Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychol. Rev.* 92, 289–316. doi: 10.1037/0033-295X.92.3.289
- Olyff, G., and Bazan, A. (2023). People solve rebuses unwittingly – both forward and backward: Empirical evidence for the mental effectiveness of the signifier. *Front. Hum. Neurosci.* 16:965183. doi: 10.3389/fnhum.2022.965183
- Paget, V., Consoli, S. M., and Carton, S. (2010). Traduction et validation française du questionnaire de répression de Weinberger. *Ann. Méd. Psychol.* 168, 593–601. doi: 10.1016/j.amp.2009.09.021
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y
- Rapaport, D. (1951). *Organization and Pathology of Thought*. New York, NY: Columbia University Press. doi: 10.1037/10584-000
- Ringel, P. Z. (1999). *Effects of Optimism and Repressive Coping on Self-Report Versus Behavioral Outcomes*. [Doctoral dissertation]. Bethlehem, PA: Lehigh University.
- Rossi, M., and Peter-Defare, E. (1998). *Les Lapsus Ou Comment Notre Fourche a Langué*. Paris: Presses Universitaires de France.
- Saraga, M., and Gasser, J. (2005). Épreuve de réalité et psychose chez Freud. *Psychothérapies* 25, 109–115.
- Scano, G. P. (2007). Who defends itself from what? Toward a reformulation of the concept of defense. *Int. Forum Psychoanal.* 16, 140–151. doi: 10.1080/08037060701302642
- Schneider, W., and Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychol. Rev.* 84, 1–66. doi: 10.1037/0033-295X.84.1.1
- Severens, E., Janssens, I., Kühn, S., Brass, M., and Hartsuiker, R. J. (2011). When the brain tames the tongue: covert editing of inappropriate language. *Psychophysiology* 48, 1252–1257. doi: 10.1111/j.1469-8986.2011.01190.x
- Severens, E., Kühn, S., Hartsuiker, R. J., and Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Soc. Cogn. Affect. Neurosci.* 7, 431–435. doi: 10.1093/scan/nsr030
- Shevrin, H. (1992). “Subliminal perception, memory, and consciousness: cognitive and dynamic perspectives,” in *Perception without awareness: Cognitive, Clinical, and Social Perspectives*, eds R. F. Bornstein and T. S. Pittman (New York, NY: Guilford Press), 123–142.
- Smith, E., and Medin, D. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674866270
- Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Tournois, J., Mesnil, F., and Kop, J. L. (2000). Autoduperie et hétéroduperie: un instrument de mesure de la désirabilité sociale. *Rev. Eur. Psychol. Appl.* 50, 219–232.
- Vézina, J. (1989). *Traduction de l'Échelle de Conformité Sociale [Unpublished manuscript]*. Québec: Université Laval.
- Wagner-Altendorf, T. A., Gottschlich, C., Robert, C., Cirkel, A., Heldmann, M., and Münte, T. F. (2020). The suppression of taboo word spoonerisms is associated with altered medial frontal negativity: an ERP study. *Front. Hum. Neurosci.* 14:368. doi: 10.3389/fnhum.2020.00368
- Weinberger, D. A. (1990). “The construct validity of the repressive coping style,” in *Repression and Dissociation*, ed. J. L. Singer (Chicago, IL: University of Chicago Press), 337–386.
- Weinberger, D. A., and Davidson, M. N. (1994). Styles of inhibiting emotional expression: distinguishing repressive coping from impression management. *J. Pers.* 62, 587–613. doi: 10.1111/j.1467-6494.1994.tb00310.x
- Weinberger, D. A., Schwartz, G. E., and Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: psychometric patterns and behavioral and physiological responses to stress. *J. Abnorm. Psychol.* 88, 369–380. doi: 10.1037/0021-843X.88.4.369
- White, K. K., Abrams, L., Koehler, S. M., and Collins, R. J. (2017). Lions, tigers, and bears, oh sh!t: semantics versus tabooeness in speech production. *Psychon. Bull. Rev.* 24, 489–495. doi: 10.3758/s13423-016-1084-8
- Xue, G., Aron, A. R., and Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebr. Cortex* 18, 1923–1932. doi: 10.1093/cercor/bhm220
- Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: a Monte-Carlo study. *Hacettepe Univ. J. Educ.* 35, 397–405.

Frontiers in Human Neuroscience

Bridges neuroscience and psychology to
understand the human brain

The second most-cited journal in the field of
psychology, that bridges research in psychology
and neuroscience to advance our understanding
of the human brain in both healthy and diseased
states.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact



Frontiers in Human Neuroscience

