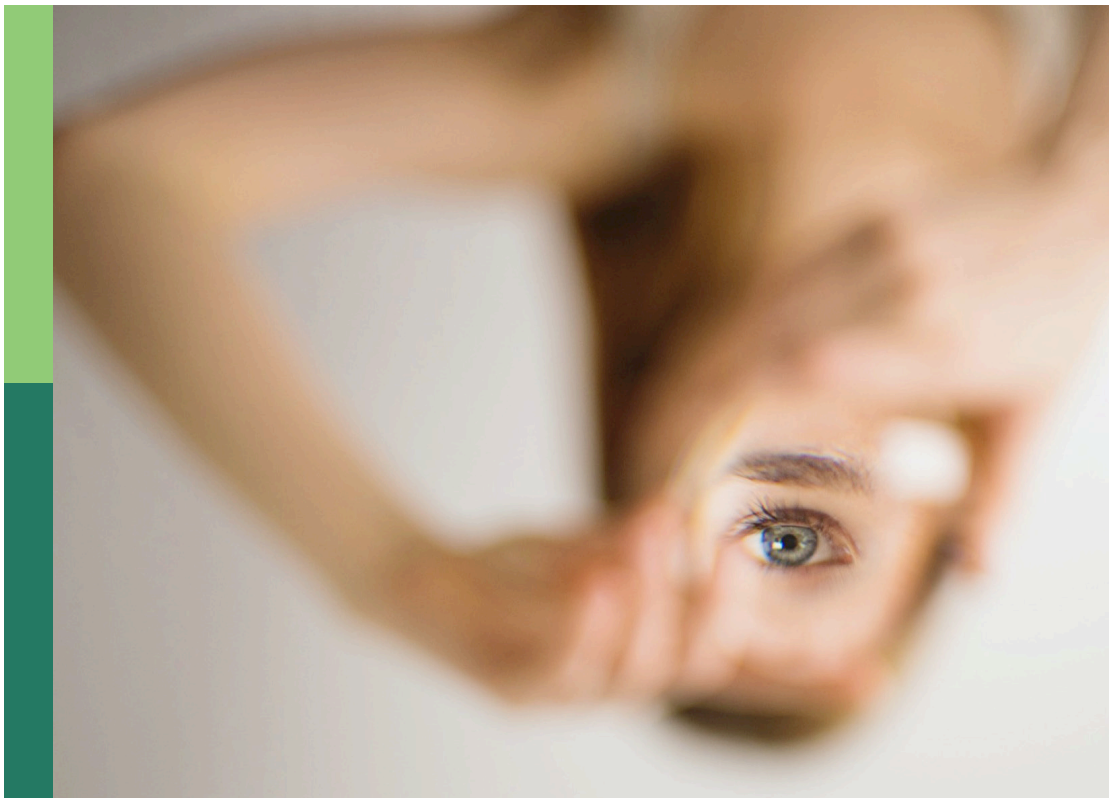
frontiers | Research Topics

Stance-taking in embodied and virtual interaction

Edited by
Geert Brône, Kurt Feyaerts, Elisabeth Zima and Silva H. Ladewig

Published in
Frontiers in Psychology



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 978-2-8325-6651-0
DOI 10.3389/978-2-8325-6651-0

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Stance-taking in embodied and virtual interaction

Topic editors
Geert Brône — KU Leuven, Belgium
Kurt Feyaerts — KU Leuven, Belgium
Elisabeth Zima — University of Freiburg, Germany
Silva H. Ladewig — University of Göttingen, Germany

Citation
Brône, G., Feyaerts, K., Zima, E., Ladewig, S. H., eds. (2025). *Stance-taking in embodied and virtual interaction*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6651-0

Table of contents

05 Editorial: Stance-taking in embodied and virtual interaction
Elisabeth Zima, Geert Brône, Kurt Feyaerts and Silva Ladewig

08 Mocking enactments: a case study of multimodal stance-stacking
Clarissa de Vries, Fien Andries and Katharina Meissl

29 Variable embodiment of stance-taking and footing in simultaneous interpreting
Alan Cienki

41 What is in a look? The accountability of gaze in trajectories to conflict
Rebecca Clift

53 Affectivity as stance: multimodal stance-taking in audiovisual documentations of Polish and German parliamentary debates
Cornelia Müller, Maciej Karpiński, Clara Kindler-Mathôt, Katarzyna Klessa, Ewa Jarmołowicz-Nowikow, Jana Katharina Junge, Katerina Papadopoulou and Brygida Sawicka-Stępińska

74 The effect of gesture expressivity on emotional resonance in storytelling interaction
Christoph Rühlemann and James Trujillo

91 Intersubjectivity and co-constructed framings: students' role-play talks in online English-speaking sessions
Leslie Huishan Li

99 Interactional practices accomplished by index-finger pointing directed at the addressee in Hebrew face-to-face interaction
Anna Inbar

118 Understanding perspective-taking in multiparty conversations: insights from Mandarin nouns
Xiaobei Zheng and Chao Sun

131 Embodied sharpness: exploring the slicing gesture in political talk shows
Silva H. Ladewig

146 Mobile stance-taking in nature: an exploration of gaze patterns during assessments of objects in nature
Barbara Laner

163 Embodiment and simulated interaction in online stance expression
Barbara Dancygier and Lieven Vandelanotte

178 Management of divergent stances as a resource to maintain progressivity and social relationships
Aija Logren, Sakari Ilomäki and Johanna Ruusuvuori

194 "That was really a productive segment, wasn't it?" Nonverbal markers of humor in the American presidential debates of 2016 and 2020
Sabina Tabacaru

203 On stance-taking with one-sided vs. two-sided shoulder lifts in German talk-in-interaction
Emma Betz and Alexandra Gubina

231 Embodying togetherness while taking divergent stances. Romantic couples' multimodal positioning practices while performing "we-stories"
Stefan Pfänder and Caroline Pfänder

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 15 July 2025
DOI 10.3389/fpsyg.2025.1636418

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Elisabeth Zima
✉ elisabeth.zima@germanistik.uni-freiburg.de

RECEIVED 27 May 2025
ACCEPTED 03 June 2025
PUBLISHED 15 July 2025

CITATION
Zima E, Brône G, Feyaerts K and Ladewig S (2025) Editorial: Stance-taking in embodied and virtual interaction.
Front. Psychol. 16:1636418.
doi: 10.3389/fpsyg.2025.1636418

COPYRIGHT
© 2025 Zima, Brône, Feyaerts and Ladewig. 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: Stance-taking in embodied and virtual interaction

Elisabeth Zima[1]*, Geert Brône[2], Kurt Feyaerts[2] and Silva Ladewig[3]

[1]Department of German Studies, University of Freiburg, Freiburg, Germany, [2]Department of Linguistics, University of Leuven, Leuven, Belgium, [3]Sign Lab, Georg-August-Universitat Gottingen, Gottingen, Germany

KEYWORDS

stance-taking, multimodality, embodied interaction, virtual interaction, gesture, multimodal packages, cognitive underpinnings

Editorial on the Research Topic
Stance-taking in embodied and virtual interaction

A plethora of conversation-analytic, discourse-functional, cognitive and psycholinguistic studies from recent years emphasize the central role of stance-taking in social interaction: "Whenever we engage in interaction, we are taking stances: there is never a time out from the social action of taking stances and adopting positions" (Du Bois and Kärkkäinen, 2012, p. 438). This act of positioning can involve three different axes, which are often considered to represent different types of stance-taking: epistemic, affective, and deontic. While epistemic stance-taking is understood as "the stance that a participant in an interaction takes with their turn in relation to a certain object of knowledge" (Deppermann, 2018, p. 121, our translation), e.g. by expressing (un)certainty, deontic stances refer to judgements about the necessity or desirability of actions, events, suggestions or similar *stance objects* (Du Bois, 2007), e.g. in cases of invitations or orders. Finally, the category of affective stances includes attitudinal and evaluative expressions which, according to Ochs (1996, p. 104), express "mood, attitude, feeling, and disposition, as well as degrees of emotional intensity vis-à-vis some focus of concern".

Following Du Bois (2007) and Jaffe (2009), stance-taking does not merely reside in the expression of one's own feelings and opinions but crucially serves to position oneself and others and thus to negotiate the interpersonal relationship by either expressing (dis-)alignment with other participants and their stances. While aligning stances serve the purpose of creating or emphasizing commonalities and thus enforcing and displaying togetherness (*interpersonal alignment*, Du Bois, 2007), disalignment may serve the purpose of setting oneself apart from others. This is reminiscent of Stivers' concept of affiliation, which she defines as the actions by which "the hearer displays support of and endorses the teller's [or interactant's, our addition] conveyed stance" (Stivers, 2008, p. 35). In contrast, disaffiliative stances express an attitude that diverges from the stance expressed by interaction partners (also Steensig and Drew, 2008; Drew and Walker, 2009 and especially Stivers, 2010, 2020). As social actions, however, affiliation and disaffiliation are not to be understood as dichotomous categories, but as two end points of a continuum (cf. Thompson et al., 2015, p. 176–178 and their analysis of stronger and weaker affiliative and disaffiliative evaluations). Similarly, the triadic relationship between stance(-taking) subjects and objects (cf. "the stance triangle", Du Bois, 2007) is not a static construct, but fundamentally dynamic in nature and subject to interactional negotiation.

Research paradigms such as Conversation Analysis, Interactional Linguistics and psycholinguistics have a long track record in the study of stance-taking as a socially contextualized and recognized interpersonal phenomenon, focusing on the lexical and grammatical resources that language users have at their disposal to communicate stance. Among these linguistic resources, special attention has gone to *pre-positioned elements* (Auer and Lindström, 2016) such as *well, okay, I don't know* etc. (Auer and Uhmann, 1982; Pomerantz, 1984; Szczepek Reed, 2015; Pekarek Doehler et al., 2021), but also to concessive constructions, modal verb constructions, and hedges. These verbal means to express stance have been studied extensively in different communicative settings (from spontaneous face-to-face communication to institutional and mediated forms of interaction), from different disciplinary angles (Interactional Linguistics, Ethnomethodology, Cognitive Psychology, HCI Research, etc.) and using different empirical methods (from controlled experiments to qualitative and quantitative corpus analysis).

However, the expression of stance - both in *face-to-face* and in virtual communication - is by no means limited to verbal means. Recent studies have explored the ways in which embodied resources such as hand gestures (see also the contributions to this Topic by Cienki, Inbar, Ladewig), body movements (Betz and Gubina), body posture, facial expressions and eye gaze (de Vries et al.; and Laner, and Clift in this Topic) play a role in the expression of stance (see Andries et al., 2023 for a systematic literature review). The goal of this Research Topic is to highlight this recent trend in the study of stance-taking and to present a state-of-the-art collection of original research that zooms in on either the co-occurrence of (i.e., "multimodal packages", Mondada, 2014) or interdependence between different semiotic resources (i.e. the sequential relationship within or across speakers).

The contributions to this Topic take a broad perspective on multimodal interaction. They unite a variety of methodological approaches, ranging from the study of naturally occurring interactions to carefully controlled experiments that tap into the cognitive underpinnings of multimodal stance-taking. Apart from the basic question on how stance is multimodally construed and negotiated in spoken and signed face-to-face interaction, the Research Topic also explores the strategies that interlocutors employ to express stance in mediated forms of interaction. It presents original research, with a particular focus on empirical studies that adopt a decidedly multimodal approach to this phenomenon.

Except for the contributions by Cienki, who focuses on stance-taking in gestures by simultaneous interpreters, and Dancygier and Vandelanotte, who study memes in online discourse, all articles in this Topic are concerned with the interactional negotiation of stance and its sequential embeddedness in different interactional contexts. These range from various forms of face-to-face interactions, including interactions in the less-studied side-by-side formations (Pfänder and Pfänder, Laner), to educational settings (Li), neurological consultations (Logren et al.), political debates (Tabacaru), and parliamentary discourse (Müller et al.) They cover the whole spectrum from affiliative to disaffiliative stance-taking (Stivers, 2008), focusing on e.g. strategies to enhance intersubjectivity (Li), humor and mockery (Tabacaru, de Vries et al.), the management of conflict (Clift) and divergent stances (Logren et al., Pfänder and Pfänder).

More specifically, the contributions by Inbar, Ladewig, Pfänder and Pfänder, Tabacaru and de Vries et al. explore how resources from different semiotic channels are combined to form recurrent multimodal packages (Mondada, 2014). Dancygier and Vandelanotte, Ladewig, Müller et al., Pfänder and Pfänder, Rühlemann and Trujillo focus on the interplay between the local, situational dimensions of stance-taking and its effects in terms of the negotiation of interpersonal relationships, emotion work and identity construction. The studies by Dancygier and Vandelanotte, Ladewig, and Li are furthermore concerned with the impact of medium—or in Laner's case the spatial formation and the interaction with the surrounding nature—on the strategies that interlocutors employ to express stance. Finally, the studies by Cienki, Rühlemann and Trujillo as well as Zheng and Sun broaden our perspective on the interactional phenomenon of stance-taking by inquiring into the cognitive processes underpinning stance-taking in interaction with Cienki also focusing on cross-linguistic differences.

In sum, this Research Topic is the first collection of original work that is specifically devoted to the study of how different semiotic resources contribute to stance-taking activities in a wide array of (interactional) contexts. What the 15 contributions to this Research Topic show, is that the topic of multimodal stance-taking can and must be approached from various angles, each of which adds to a better understanding of what is one of the most fundamental social activities realized in/through language and embodied behavior. While qualitative studies in the tradition of Conversation Analysis and Interactional Linguistics provide us with relevant insights into the sequential organization and social dynamics of multimodal stance-taking, the more quantitatively oriented experimental and corpus-based studies help us in identifying patterns and relevant factors that drive the expression and negotiation of stance. It was the explicit intention of this Research Topic to unite these different perspectives and thus to set the agenda for future research on the multifaceted nature of stance-taking in embodied and virtual interaction.

Author contributions

EZ: Writing – review & editing, Writing – original draft. GB: Writing – review & editing, Writing – original draft. KF: Writing – original draft, Writing – review & editing. SL: Writing – review & editing, Writing – original draft.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). Multimodal stance-taking in interaction—A systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Auer, P., and Lindström, J. (2016). Left/right asymmetries and the grammar of pre- vs. post-positioning in German and Swedish talk-in-interaction. *Lang. Sci.* 56, 68–92. doi: 10.1016/j.langsci.2016.03.001

Auer, P., and Uhmann, S. (1982). Aspekte der konversationellen Organisation von Bewertungen. *Deutsche Sprache* 1, 1–32.

Deppermann, A. (2018). "Wissen im Gespräch," in *Handbuch Text und Gespräch*, ed. K. Birkner, and N. Janich (Berlin: De Gruyter), 104–142.

Drew, P., and Walker, T. (2009). Going too far: Complaining, escalating and disaffiliation. *J. Pragmat.* 41, 2400–2414. doi: 10.1016/j.pragma.2008.09.046

Du Bois, J. W. (2007). "The stance triangle," in *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. Englebretson, R. (Amsterdam: Benjamins), 139–182.

Du Bois, J. W., and Kärkkäinen, E. (2012). Taking a stance on emotion: affect, sequence, and intersubjectivity in dialogic interaction. *Text Talk* 32, 433–451. doi: 10.1515/text-2012-0021

Jaffe, A. M. (2009). *Stance: Sociolinguistic Perspectives*. Oxford, NY: Oxford University Press.

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *J. Pragmat.* 65, 137–156. doi: 10.1016/j.pragma.2014.04.004

Ochs, E. (1996). "Linguistic resources for socializing humanity," in *Rethinking Linguistic Relativity*, ed. Gumperz, J., Levinson, S. (Cambridge: Cambridge University Press), 407–437.

Pekarek Doehler, S., Polak-Yitzhaki, H., Li, X., Stoenica, I. M., Havlík, M., and Keevallik, L. (2021). Multimodal assemblies for prefacing a dispreferred response: a cross-linguistic analysis. *Front. Psychol.* 12:689275. doi: 10.3389/fpsyg.2021.689275

Pomerantz, A. (1984). "Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes," in *Structures of Social Action: Studies in Conversation Analysis,* ed. M. Atkinson, and J. Heritage (Cambridge: Cambridge University Press), 57–101.

Steensig, J., and Drew, P. (2008). Introduction: questioning and affiliation/disaffiliation in interaction. *Discourse Stud.* 10, 5–15. doi: 10.1177/1461445607085581

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stivers, T. (2010). An overview of the question–response system in American English conversation. *J. Pragmat.* 42, 2772–2781. doi: 10.1016/j.pragma.2010.04.011

Stivers, T. (2020). *The Book of Answers*. Oxford: Oxford University Press.

Szczepek Reed, B. (2015). Managing the boundary between "yes" and "but": two ways of disaffiliating with "ja aber" and "jaber." *Res. Lang. Soc. Interact.* 48, 32–57. doi: 10.1080/08351813.2015.993843

Thompson, S., Fox, B., and Couper-Kuhlen, E. (2015). *Grammar in Everyday Talk: Building Responsive Action*. Cambridge: CUP.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 02 April 2024
DOI 10.3389/fpsyg.2024.1379593

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Omid Khatin-Zadeh,
University of Electronic Science and Technology of China, China
Isabella Poggi,
Roma Tre University, Italy

*CORRESPONDENCE
Fien Andries
✉ fien.andries@kuleuven.be

†These authors have contributed equally to this work and share first authorship

RECEIVED 31 January 2024
ACCEPTED 26 February 2024
PUBLISHED 02 April 2024

CITATION
de Vries C, Andries F and Meissl K (2024) Mocking enactments: a case study of multimodal stance-stacking.
Front. Psychol. 15:1379593.
doi: 10.3389/fpsyg.2024.1379593

COPYRIGHT
© 2024 de Vries, Andries and Meissl. 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.

Mocking enactments: a case study of multimodal stance-stacking

Clarissa de Vries[1†], Fien Andries[2*†] and Katharina Meissl[1†]

[1]Department of Linguistics, Faculty of Arts, KU Leuven, Leuven, Belgium, [2]Department of Linguistics, Faculty of Arts, KU Leuven, Antwerp, Belgium

Although research into multimodal stance-taking has gained momentum over the past years, the multimodal construction of so-called stacked stances has not yet received systematic attention in the literature. Mocking enactments are a prime example of such complex social actions as they are layered both interactionally and stance-related, and they rely significantly on the use of bodily visual resources, depicting rather than describing events and stances. Using Du Bois' Stance Triangle as a framework, this study investigates mocking enactments as a case study to unravel the multimodal aspects of layered stance expressions. Drawing on three data sets—music instruction in Dutch, German, and English, spontaneous face-to-face interactions among friends in Dutch, and narrations on past events in Flemish Sign Language (VGT)—this study provides a qualitative exploration of mocking enactments across different communicative settings, languages, and modalities. The study achieves three main objectives: (1) illuminating how enactments are used for mocking, (2) identifying the layers of stance-taking at play, and (3) examining the multimodal construction of mocking enactments. Our analysis reveals various different uses of enactments for mocking. Aside from enacting the target of the mockery, participants can include other characters and viewpoints, highlighting the breadth of the phenomenon under scrutiny. Second, we uncover the layered construction of stance on all axes of the Stance Triangle (evaluation, positioning, and alignment). Third, we find that mocking enactments are embedded in highly evaluative contexts, indexed by the use of bodily visual resources. Interestingly, not all mocking enactments include a multimodally exaggerated depiction, but instead, some merely allude to an absurd hypothetical scenario. Our findings contribute to the growing body of literature on multimodal stance-taking, by showing how a nuanced interpretation of the Stance Triangle can offer a useful framework for analyzing layered stance acts.

KEYWORDS

mocking, stance, enactment, depiction, multimodality, Dutch, German, Flemish Sign Language

1 Introduction

Mocking is a form of playful exchange, in which stance takes center stage. During mocking, participants express a stance on a serious layer that can be heightened, diminished, or inverted on a non-serious layer. In other words, they stack stances onto each other (Dancygier, 2012; Andries et al., 2022). One resource that is eminently suitable for the expression of stances on different interactional layers is enactment as it allows interactants "to

'construct' actions and dialog in order to 'show' characters, events, and points of view" as they combine "bodily movements, postures, and eye gaze" (Hodge and Ferrara, 2014, p. 373). In enactments, participants may represent a stance of the character they enact, i.e., showing what this character feels and thinks, what their opinion is, what they know, or present their own stance, from a narrator's point of view. Moreover, multiple stances may be combined and stacked onto each other, as is the case in mocking enactments. Consider the following example from one of the data sets of the current study. Three friends (Emma, Jana, and Alyssa) are talking about a time when Jana went away for a weekend with a group of scouts and stayed in a house in the woods that had no keys (so that the house could not be locked). Jana enacts the landlady of the house, who did not consider this to be a problem. Accompanied by multiple shoulder shrugs, palm up open hand gestures, and head shakes, she says "oh gosh, well, throughout the years, all the keys just got lost." Thereby, Jana mocks the indifferent attitude of the landlady toward this issue.

As illustrated by this example, both in enactments and during mocking, participants leverage an array of both lexical and non-lexical resources across various modalities. However, the research on multimodal aspects of how enactments are used in layered stance expressions remains limited so far. In this study, mocking enactments serve as a case study to unravel multimodal dynamics of stance-stacking and the underlying mechanisms at play. We zoom in on mocking enactments as a case of stance-taking as a multimodal, polysemiotic[1] phenomenon, without a predetermined hierarchy of semiotic resources; instead, they are employed flexibly, contingent upon the interactional context. In this exploration, we use the Stance Triangle (Du Bois, 2007) as a conceptual framework and consider three dimensions of stance: evaluation of the stance object, positioning of stance subjects, and their alignment. Our focus is on dissecting and understanding the intricate dynamics of mocking enactments, shedding light on the nuanced layers of stance inherent in these communicative acts.

The current exploratory study pursues three primary objectives. First, we aim to illuminate the ways in which enactments are used for mocking. Second, we identify the layers of stance-taking operative across all components of the Stance Triangle. Third, we examine the multimodal construction of these enactments and layers of stance. To achieve these aims, the study draws on corpora from three settings: (1) music instruction in Dutch (Schrooten and Feyaerts, 2020), as well as German and English (MuTh, 2021), (2) spontaneous face-to-face interactions among friends in Dutch (Brône and Oben, 2015; de Vries et al., n.d.), (3) narrations of past events from the corpus Flemish Sign Language (VGT, Van Herreweghe et al., 2015). The combination of three data sets allows for a broad and yet qualitative approach to mocking enactments in different communicative settings as well as languages in different modalities and from different communities.

In Section 2, we delve deeper into the multimodal analysis of stance-taking, mocking, and its relation to stance-stacking, as well as enactment as a polyphonic communicative device. Section 3 contains our methodological approach, along with a presentation of the three different data sets used in the study. The analysis is presented in Section 4, starting with an emphasis on the evaluation process of the Stance Triangle. Subsequently, we shift our focus to the theme of positioning, and finally, we scrutinize alignment. The study's findings are synthesized and discussed in Section 5. A conclusion, encapsulating the key insights and implications, is presented in Section 6.

2 Theoretical background and rationale

2.1 Multimodal stance-taking

One of the primary functions of language and interaction more broadly is to communicate how we feel about the world around us. Throughout our daily lives, the negotiation of our attitudes is ubiquitous: "do not eat that"; "did you like the show?"; and "I do not know how this works," are just a few everyday examples of the linguistic negotiation of stance in different dimensions (in this case, deontic, affective, and epistemic, respectively).

In the current study, we draw on the influential analytical framework of the Stance Triangle, by Du Bois (2007), as it presents a broad view on stance-taking and brings together cognitive and interactional dimensions. Three questions at the core of the framework direct us to the components of stance-taking: "*(1) Who is the stancetaker? (2) What is the object of stance? (3) What stance is the stancetaker responding to?*" (Du Bois, 2007, p. 146). These questions connect stance subjects and stance objects through three processes: evaluation, positioning, and alignment. The processes and their relation to stance subjects and objects can be visualized in a triangle (see Figure 1).

As can be seen in Figure 1, the result is a dynamic and intersubjective process of stance negotiation, which Du Bois captures in the following definition: "Stance is a public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self and

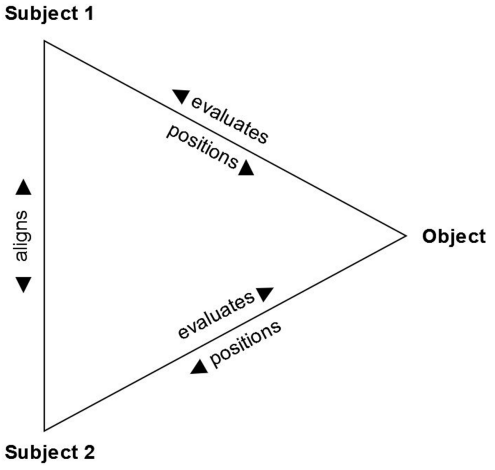


FIGURE 1
The stance triangle (adapted from Du Bois, 2007, p. 163).

1 Within this study, we view stance, enactment, and mocking as multimodal and polysemiotic phenomena as they may involve different perceptual modalities and semiotic systems. For an elaborate discussion of the use of the terms polysemiotic and multimodal, we refer the reader to Zlatev (2019).

others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field" (Du Bois, 2007, p. 163). More recently, Iwasaki (2022) argued for an elaboration of the Stance Triangle so as to take into account the evolvement of stances across longer stretches of time. While the Stance Triangle can be considered as a 'snapshot' of the stance at a given moment in interaction, Iwasaki (2022, p. 15) highlights that "all aspects of the stance activity are in constant motion and emerge across several turns along the temporal trajectory during an interaction".

Furthermore, although the Stance Triangle is modality agnostic in theory, traditionally, the focus of stance-taking research has been on lexico-grammatical and to a lesser extent prosodic means (Biber and Finegan, 1989; Couper-Kuhlen, 2012; Wang et al., 2022). Over the past years, interest in multimodal research on stance-taking has increased (Andries et al., 2023). Research in this domain highlighted that the whole body can be used to take a stance, with individual semiotic resources operating flexibly depending on the interactional needs. The compound shrug is a prototypical example, as it consists of a varying constellation including raised shoulders, palm up gestures, head tilts, and raised eyebrows, and can express multiple forms of distancing, such as obviousness, indifference, or incapacity (Debras, 2017), as we saw in the example above with the lost keys. Furthermore, the affordances of bodily visual resources are highly suitable to monitor and negotiate the temporal development of stance as they allow for ensuring fit responses (Pillet-Shore, 2020), negotiating alignment (Iwasaki, 2015), and synchronizing the expression of stance (Pfänder and Couper-Kuhlen, 2019), without disrupting the flow of conversation. Finally, and of relevance for the current study, multiple semiotic resources can be used simultaneously not only in function of one holistic "gestalt" (Mondada, 2014) such as the shrug but also in order to allow signers/speakers to present multiple stances or viewpoints at the same time (Kärkkäinen, 2012; Soulaimani, 2018; Vandenitte, 2022).

Although there is considerable evidence for the inherent multimodal nature of stance-taking (Andries et al., 2023), there are still many open questions. One aspect that has not received systematic attention yet concerns the construction of so-called stance-stacks (Dancygier, 2012) in which multiple stances are communicated simultaneously. The current study therefore aims to scrutinize this phenomenon, by looking at a case study of layered stance-taking that has been surprisingly understudied: the use of mocking enactments. In what follows, we will briefly review the literature on mockery (Section 2.2) and enactments and depictions (Section 2.3) before presenting the research questions and aims of the current study.

2.2 Mocking as stance-stacking

Mocking concerns a playfully constructed negative stance toward something of relevance to a target. As the term itself implies, mocking involves some kind of pretense or imitation. As such, it can be considered a *staged communicative act* (Clark, 1996). Staged communicative acts involve a layered construction of actions, in which language users engage in a *joint pretense* (Clark and Gerrig, 1984). A signer/speaker addresses an addressee A, while at the same time pretending to be another signer/speaker S′, addressing another addressee A′ (Clark and Gerrig, 1984, p. 122). The contrast between the demonstrated and actual situation gives rise to the mocking, which the addressee is expected to recognize and appreciate. For instance, in the example with the missing keys from the introduction, Jana enacts the landlady's stance and thereby pretends to be indifferent about the lost keys. Simultaneously, Jana's own stance—that this is unsafe—becomes apparent, resulting in a stance-stack with a mocking quality (Janzen, 2022).

Of particular relevance in the context of stance-taking is the fact that mocking is directed at a target, which can be the signer/speaker themselves, another participant in the interaction, or a non-co-present third party. Depending on the target of the mockery and the interactional setting, mocking can fulfill a wide range of functions, including bringing shared amusement (Yu, 2013), sanctioning transgressions (Drew, 1987), managing errors in instruction (Poggi, 2015), and building relationships (Boxer and Cortés-Conde, 1997). Arguably, the non-serious layer allows participants to deal with potentially sensitive issues. Or, in the words of Holt (2013, p. 89): "Introducing non-seriousness does not mean that the serious sequential implications are completely swept away. Rather, it enables an intertwining of strands as serious matters are dealt with more and less seriously, allowing for a more delicate and implicit touch". For instance, during orchestra rehearsals, conductors can use mocking to point to errors in the musicians' performance, while maintaining a friendly atmosphere (Poggi, 2015).

Crucially, mocking is constructed and negotiated in interaction, which can be quite complex given its layered nature. Bodily visual practices have often been described to play a large role in this regard (Clark, 1996, p. 370), and in what follows we will present an overview of research into the multimodal construction of mocking. As research on multimodal aspects of mockery itself is rather limited, we also draw on work on related forms of staged communicative acts, including irony and humor.

The intuitively most obvious bodily visual practices used to construct non-seriousness are laughter and smiling. Turn-final laughter, for instance, can function to frame an utterance as humorous or non-serious (Holt, 2013), as well as to invite other participants to laugh along and thereby affiliate with an interlocutor (Jefferson et al., 1987). The closely related resource of smiling has traditionally also been associated with non-seriousness. Recent studies confirm that an increase in smiling intensity (by both speakers and addressees) is associated with the use of humor (Gironzetti, 2022) and irony (González-Fuente et al., 2015). However, the relationship of smiling and laughter with affiliation and non-seriousness is not straightforward. Laughter occurs both in the context of 'laughing with' and 'laughing at' (Glenn, 2003), or both at the same time (Clift, 2016), and smiling, similarly, is also found to be associated with embarrassment or nervosity (Ambadar et al., 2009). Furthermore, the presence of laughter or smiling is not a prerequisite for the successful performance of humor (Priego-Valverde and Rauzy, 2023).

Other resources can be ascribed to the staging of an exaggerated scenario or otherwise highlighting the incongruence between reality and expectation. Earlier work on verbal cues in written discourse mentions the use of extreme case formulations or overstatements (Norrick, 2004) to convey humorous or ironic intent, as well as shifts in register or other formulaic markers (Burgers, 2010). The multimodal staging of mocking is often connected to its stance-related aspects (see Section 2.1 for a discussion of multimodal markers of stance). Such realizations include covering the face with the hands in response to a laughable, throwing the head back, or shaking the torso

and shoulders (Ford and Fox, 2010). Gaze aversion has frequently been mentioned as a cue to convey ironic intent by speakers (Caucci and Kreuz, 2012; González-Fuente et al., 2015; Gironzetti, 2022) and can also be linked to a stance-taking role, namely, displaying disalignment (Haddington, 2006), or in the context of self-mockery, embarrassment (Yu, 2013).

Multimodal complements for exaggeration and overstatements can be found, for example, in the case of parodies, where a "distorted imitation" is performed, "in which the target's traits are exaggerated and made grotesque" (D'Errico and Poggi, 2016, p. 4), a finding also reported in Ford and Fox's (2010) analysis of constructing laughables. The whole gamut of semiotic resources can be crucial when it comes to staging or animating characters as part of a mocking event. In what follows, we dig deeper into the design and use of enactment.

2.3 Enactment and depiction

Before we turn to the specific role of enactment for stance-taking and mocking, let us introduce the phenomenon in general. Enactment refers to "signers and speakers combining bodily movements, postures, and eye gaze to 'construct' actions and dialog in order to 'show' characters, events, and points of view" (Hodge and Ferrara, 2014, p. 373).

As such, enactments fulfill the function of *depicting* meaning, as distinguished from the other methods of communication, *describing,* and *indicating* (Clark, 1996, 2016). Clark (2016, p. 325) defines depictions as "physical scenes that people stage for others to use in imagining the scenes depicted". Whereas depiction is a broader phenomenon of iconic meaning representation, during enactment, signers/speakers map a referent onto their body "on a real-world scale" (Cormier et al., 2013, p. 370), constructing this referent's actions and/or dialogs (see Beukeleers and Vermeerbergen, 2022 for a historical overview of the terminology). Imagine the difference in options for depicting someone who is walking. One option would be using your hand, with index and middle finger pointing downward and moving alternatingly for depicting on a miniature scale with a gesture or sign. Another option would be moving your arms as if walking, where the act of walking is mapped onto the signer's/speaker's body, and therefore considered an enactment.

Not only enactment itself but also the projection and framing of enactments in a stretch of discourse is multimodal. There is a range of lexical cues that have been described for English to open slots for enactments, such as verbal indices (*like this*), quotatives (*say, ask*), particles (*just, kind of, like*), or other syntactic projections (Tannen, 1986; Hsu, 2021, p. 105; Hsu et al., 2021). Investigations of enactments in British and Australian Sign Language (BSL, Auslan) have shown that signers typically frame their enactments with lexical noun phrases and/or pointing actions, which index the subsequently enacted referent (Cormier et al., 2013; Ferrara and Johnston, 2014). In spoken interaction, prosodic shifts, such as the use of lower, or higher pitch or a louder voice and changes in vowel quality, often proceed enactments (Günthner, 1997). Additionally, in both signed and spoken conversations, shifts in eye gaze (Engberg-Pedersen, 1993; Maury-Rouan, 2011; Thompson and Suzuki, 2014) as well as body orientation (Cantarutti, 2020) or head position (Maury-Rouan, 2011) are indicators that an enactment will follow. Specific gaze patterns have been observed, for example, depending on whether the enacted scene is dialogic or not (Pfeiffer and Weiss, 2022). Furthermore, qualities of the enactment, such as specific prosody or facial expressions, are sometimes already used in the discourse leading up to the enactment and therefore contribute to its framing through a spillover effect (Cantarutti, 2020).

A body of work shows that enactment and depiction practices form a very common way of representing information (Clark, 2016; Ferrara and Hodge, 2018; Hsu et al., 2021; Beukeleers and Vermeerbergen, 2022), both in signed and spoken languages. In narratives, for example, enactment is frequently used to show how something happened or what was said and how. Viewpoints and roles can alternate quickly, shifting back and forth from show to tell. Moreover, viewpoints can be combined through body partitioning, as introduced by Dudis (2004), including or excluding the narrator. Signers/speakers thus often shift back and forth between a narrator's viewpoint and a character's viewpoint, to vividly depict an action or discourse situation from a character's viewpoint, or they combine narration with enactment.

More generally, enactments are always layered and constructed since they are the result of an interactant creatively and selectively "replaying" (Goffman, 1974) activities and (linguistic) actions. Therefore, according to Besnier (1993, p. 161) they are, on the one hand, "the representation of linguistic actions" and, on the other hand, "commentaries about these actions." This makes enactments a useful tool for expressing stances in interaction (Niemelä, 2010; Cantarutti, 2020). For instance, in the case of music instruction, sonic qualities or movements for instrument manipulation can be depicted as a part of the evaluation of performances and subsequent instruction for musicians by instructors (Meissl et al., 2022). When signers/speakers express a stance during enactment, they create a complex stance-stacking act (Andries et al., 2022; Janzen, 2022). They not only express their stance as a narrator in the here and now but also stack this upon the inherent stance of evidentiality from reporting the speech or action (Shaffer, 2012). Moreover, enactment sequences may report the stances of the enacted characters the signer/speaker depicts during a past event or discourse situation (Debras, 2015). Connected to the expression of a third party's stance, Debras (2015) has suggested extending Du Bois' triangle to a tetrad, where another stance subject is added.

Because of the inherently layered nature of enactments and the potential for enactments to express polyphonic (Günthner, 1999) stances, they are a suitable device for mocking. As Debras (2015) states, speakers can distance themselves from an absent subject's stance by enacting it. As such, multiple stances can be expressed through enactment simultaneously, e.g., by enacting a referent in an exaggerated or stereotypical way to mock them by means of a caricature of a social category attributed to the referent. Fischer and Simon (2016), for example, describe how DGS (German Sign Language) signers use exaggerated mouth gestures in negatively evaluative enactments of hearing people. However, as will be shown in Section 4, mocking can be achieved through a range of different mechanisms next to parodistic imitations. Following the recent studies of Cantarutti (2020) and Mandel (2022), we scrutinize the potential of enactments as a resource for layered stance expressions.

2.4 Research aims

The aims of the current contribution are to shed light on (1) how enactments are used for mocking; (2) which layers of stance-taking (on all components of the Stance Triangle) are at play; and (3) how these enactments (and layers of stance) are multimodally constructed.

To feed into the overarching research question on how enactment is used for mocking and which layers are construed multimodally as a result, several smaller-scale questions guided the analysis. For each example, we will scrutinize (a) who or what is enacted, (b) what the stance object is, and (c) who is the target of the mocking. Building on these questions, Section 4.2 discusses (d) what position the stance subjects assume and Section 4.3 focuses on (e) how stance subjects negotiate alignment. These questions allow for teasing apart layers on each component of the Stance Triangle.

3 Data and method

To address our research questions, we analyzed the phenomenon of mocking enactments in different interactional settings and several languages, including three spoken and one signed language.

For the **music instruction setting**, two video corpora were consulted. First of all, we used approximately 27 h of wind and brass orchestra rehearsals in Flanders, Belgium (Schrooten and Feyaerts, 2020). Five different conductors were filmed during three rehearsals each with their respective ensemble, and the language of interaction is Dutch. The second corpus contains approximately 8 h of recordings from three chamber music coaching sessions (MuTh, 2021), in which one string quartet worked with three different coaches on consecutive days. The first two sessions are in German and the third in English.

Concerning data on **spontaneous conversations between friends**, two video corpora were used, amounting to 9 h of data in total. Both corpora consist of Dutch triadic interactions. The first corpus (Brône and Oben, 2015) includes eight recordings of spontaneous interactions and eight recordings of brainstorming sessions,[2] in which participants wear head-mounted eye trackers (for an elaborate description of the corpus and setup, see Jehoul, 2019). The second corpus (de Vries et al., n.d.) contains 12 recordings of spontaneous triadic interactions between friends in a coffee bar and includes 3 camera perspectives on the faces and upper body of each participant. Participants received no instructions for these conversations.

The **narrative data in Flemish Sign Language** are dyadic conversations taken from the corpus Flemish Sign Language (Van Herreweghe et al., 2015). The data used for this project consist of 5 h of dyadic conversations of 34 signers. Two types of conversations are used: (1) free conversations without a moderator being present, during which the participants could talk about whatever they liked, and (2) guided conversations about topical past events. During those guided conversations, the participants were shown a photograph of a historical event such as 9/11, the 2004 Indian Ocean earthquake and tsunami, and the conviction of the Belgian pedophile Marc Dutroux. The signers were invited to talk about what they remember of this event, where they were when they learned what had happened, and how they experienced it.

We first scanned the different video corpora for sequences with mocking enactments. As is the case with many phenomena connected to humor or non-seriousness in interaction (Gibbs, 2000, p. 12), it is not always evident to draw strict lines between what counts as 'mocking enactment' or not. To balance out individual subjective judgments, we maintained an inter-coder negotiation process throughout the data segmentation and analysis period. As a baseline, approximately 10 cases per data set were discussed in groups of three. For the data in VGT, two deaf signers were consulted for their judgment. For the data in spoken languages, the three authors discussed the case selection, and throughout the analysis process, cases of doubt were double-checked within this team. These discussions facilitated a selection of approximately 30 cases per setting that form the basis for the current analyses. Tendencies and patterns that surfaced in these cases will be discussed on the basis of five examples in the following sections.

4 Analysis

4.1 Evaluation

In this first part of the analysis, we turn to the music instruction setting and discuss how mocking enactments can be used for purposes of evaluation during instruction. The two examples illustrate how the use of absurd imagery by a conductor (Example 1) as well as a self-deprecating enactment by the coach in a master class (Example 2) can contribute to evaluations of stance objects on different layers. For each example, we will discuss who or what is enacted, what the stance object is, and who the target of the mocking is, to then elaborate on how this feeds into the evaluation process that is central to the setting of music instruction.

As a first example, we will look at an extract taken from a recording of an orchestra rehearsal (see Supplementary material S1).[3] In this setting, there is a characteristic temporal organization of alternating sequences of play and sequences in which play is interrupted and the conductor offers feedback and instructions (cf. Stoeckl and Messner, 2021). Example 1 marks a typical instructional sequence, in which a contrast between an undesired and desired performance is expressed (Weeks, 1996; Meissl et al., 2022). We will discuss the example in two parts, initially focusing on lines 01–08 where the first mocking enactment is staged, and then turn to the rest of the sequence in lines 09–21 with subsequent enactments.

The conductor interrupts a playing sequence after the first few notes and addresses the flute section (lines 01–02). With a raised voice, the conductor utters a rhetorical question, asking the musicians what scares them off about their first note—a B-flat—and thereby localizes the problem in their previous performance: their hesitancy

2 For the brainstorms, participants received the instruction to talk about their ideal student house and bar. In practice, this resulted in spontaneous discussions of houses that participants lived in and bars that they went to.

3 For the spoken language examples, GAT2 conventions were used for the transcriptions of speech (Selting et al., 2009), and Mondada's (2018) conventions for transcribing multimodal conduct.

and a delayed onset of the first note. This question is quite affectively loaded given the prosody and the conductor's tense gesture with clenched fists in line 03. He reformulates his question and leans slightly backward (line 04), distancing himself physically from the musicians. Lines 01–04 therefore mark the first negative evaluation of the preceding performance as a stance object.

After a short pause, in line 06, the conductor sets the stage for the mocking enactment by uttering "da's zo" (*that's like*), referring to the (performance of the) first note. Typical for this setting, the formulation he uses projects information about the undesired performance. The opened syntactical slot is filled by an enactment, which can be split into two parts, the first marked by a sense of 'tension' and the second by 'release'. During the first phase, the conductor clenches his fists in front of his body with his arms angled and pressed tightly toward the torso. He frown-raises his eyebrows (Nota et al., 2021) and pulls the corners of his mouth down with tense closed lips. Next, he tilts his head back and looks up. This very tense bodily display is accompanied by two vocalized sounds with a very pressed and tense quality—first a restricted "m" and then a slightly louder "bi" (Figure 2 - Image 1). The second phase of the enactment is initiated with an eyebrow raise, a short flash of puffed cheeks, and subsequently the conductor opens both hands to a cup shape, as if holding an object. This goes together with a sound such as "bwe" and after the hands are already opened, another "plonk" during which the conductor looks at the orchestra (Figure 2 - Image 2). The interplay of visual resources and the sounds with a 'plopping' quality results in an expression of tension release. The associations that this complex display may invoke, namely, a chicken laying an egg, are immediately verbalized in line 07. When producing this utterance, the conductor also loosens tension in his shoulders and arms, still contributing to the overall impression of 'tension release.'

```
01   C4     ♫ja maar-
            yes but
     mus    ♫some continue playing-->
02   C4     %(0.4)%
     c4     %claps hands once%
03   C4     %wat (0.2) wat +SCHRIkt er u af%%in die si MOL,
             what scares you off about that b flat,
     c4     %clenches fists ifo torso------%%drops arms--->
                            +shifts weight and leans back-->
04   C4     dat+ vraag ik me af.♫%+
            that I'm asking myself
     c4     ---------------------->%
     c4     ---+shifts weight forward+
     mus    ---------------------->♫
05   C4     (0.4)
06   C4     da's zo+  %&§(0.4)% *     m bi#     %*&§ &§(0.3)§ %    %bwe*    &plo#nk%
            that's like
     c4     +straightens
                 torso+
     c4                %.......%clenches fists,                %...%opens hands---%
                          presses arms to torso%
     c4                 &frown-raises eyebrows----&  &raises eyebrows-----&
     c4                  §pulls mouth corners down-§  §puffs cheeks§
     c4                         *tilts head back--------------------*
     fig                              #1                                  #2
```

Image 1 Image 2

```
07   C4     %en daar ligt% een ei;
            and there is an egg
     c4     %releases
              arm tension%
08   C4     ♫(1.0)♫
     mus    ♫((laughter from some musicians))♫
```

FIGURE 2
Example 1.

We therefore see, that in the enactment (line 06) and subsequent description (line 07) of the undesired performance, the conductor is referring to the musicians' performance on the interactional base layer. However, he does so through the addition of absurd imagery on a pretense layer, comparing the onset of a note with laying an egg. Put into perspective with the Stance Triangle, another stance object is introduced: A chicken laying an egg, attached to which comes a whole set of associations with chickens and their behavior. Retrospectively, that semantic domain also fits the feeling of "being scared off," which the conductor attributes to the musicians in line 03. The chicken laying an egg is mapped onto the conductor's body, who is also an evaluating stance subject. Through the use of this enactment, the conductor evaluates the imagery of a chicken laying an egg as negative, which in turn serves as the negative evaluation of the musicians' performance. This layering results in a mockery of the musicians' previous performance of the first note. During a 1.0-s pause in line 08, there is audible laughter from some members of the orchestra, showing that the enactment has caused amusement.

Now, we will turn to lines 09–21 of the example to see how the conductors' evaluation and instruction progress. In lines 09–11, the conductor depicts and describes the desired performance, highlighting the initiation of the first note (Figure 3 - Image 3). The contrastive conjunction "but" (line 12) projects the return of focus to the undesired performance, which is continued in line 14 with a slot opener for another mocking enactment. This shares a range of qualities with the first part of the previous egg-laying enactment in line 06. The conductor clenches his hands in a claw handshape, presses his arms to his torso as before, and pulls up his shoulders. His head is tilted right and pulled down, resulting in a compression of his torso. The tension is also visible in his face, with a frown-raise and pressed as well as puffed lips. Through these tense lips, some suppressed sounds are audible. Thus, the notion of tension and effort is reinvoked (Figure 3 - Image 4). The release, however, only follows with the utterance "and it does not come" (line 15), a construction that expresses both the desired result—the note coming out—and the fact that this had not happened in the previous performance (Figure 3 - Image 5). Thereby, the conductor highlights the incongruence between his expectations and the reality of the musicians' actions. What follows this extract is the repetition of the desired performance qualities (Figure 3 - Image 6) and the initiation of another playing sequence in line 21.

Summing up, Example 1 features two mocking enactments, in which a conductor offers parodistic imitations of the musicians' previous performance, a common tool in this setting to highlight undesired qualities of performances (Poggi, 2015). We have shown that this results in the layering of the stance object and evaluation through exaggeration and absurd imagery.

While such parodies of co-present participants are probably the most common format in which mocking enactments appear in the setting of music instruction, we will now turn to an instance in which a mocking enactment is used as a form of self-deprecation based on invoking an absurd scenario (see Supplementary material S2). A different kind of evaluative layering is brought about by coach Stefan Gottfried (henceforth SG) in a string quartet master class. The ensemble members are—seated from left to right—Marie-Therese Schwöllinger (MTS), Alexandra Moser (AM), Anuschka Cidlinsky (AC), and Oscar Hagen (OH).

During the session preceding Example 2, coach SG has suggested to use *cesuras* as expressive elements. A cesura is an interruption between notes that can function in a way similar to a comma in a spoken sentence, resulting in a pause between two phrases. After several sequences of play and interruptions in between, SG returns to the topic of cesuras in Example 2, referring to a specific instantiation in the piece (line 03). He evaluates the realization of the cesura positively as 'great', smiles, and tilts his torso to the right and slightly forward (Figure 4 - Image 7), which can be interpreted as a hedging device projecting the subsequent modification of his evaluation in line 04. SG utters "die is mir jetzt" (*that is now for me*), stressing "mir" (*for me*) and touching his chest as a deictic gesture toward himself. Before this sentence is finished, however, an adverbial contrastive subclause is inserted, which also contains the mocking enactment. SG says "wo ich doch so" (*while I am so much*) (line 04) and laughs for about 1.0 s (line 05), during which he lifts his two hands up as if holding up a sign, mapping the action onto his body. While holding his hands in the same way above head height and smiling (Figure 4 - Image 8), he utters "für zäsuren werbe" (*advertizing cesuras*), completing the subclause that as a whole translates roughly to 'while I am advertising cesuras so much' and therefore refers back to the instructions he has given before this extract. Completing the main clause that was started in line 04, SG finishes his evaluation of the performed cesura in line 07 (Figure 5) by saying that it was a bit too much for him there. Through the mocking enactment, in which SG frames his repeated instructions as an advertisement for an artistic means of expression, he positions his previous instructions as the stance object and himself as the target of the mockery. SG marks his own previous actions as laughable, which is reciprocated by the musicians through smiling (lines 05–08).

The stance object therefore evolves from the musicians' realization of the cesura to the instructions that SG has previously given. Note that in contrast to the mocking enactments in Example 1, the enactment by SG here does not draw on means of exaggeration in its multimodal orchestration. It is rather the absurdity that lies in the analogy through which advertisement and holding up a sign for promotion is mapped onto the action of instructing in a musical setting. Through this absurd analogy, another level of mocking is achieved which lies in the incongruence between his previous and current stance toward cesuras. This is highlighted by the use of the inserted contrastive subclause (lines 04–05) in which the enactment is embedded. While promoting cesuras before, and therefore influencing the musicians' performance now up for evaluation, SG deems the cesura too much at that specific point (line 06). He acknowledges that incongruence through self-deprecation which allows him to take responsibility for the musicians' performance of said cesura and results in smiles from the members of the ensemble.

In line 07, the first violinist (MTS) repeats the last words of SG's ongoing turn in overlap, showing uptake of that evaluation, and gives a confirming 'mhm' in line 09. Following that, SG goes on to explain, what the desired rendition of the fragment in question would be (lines 11–13), in contrast to the unwanted interruption through the cesura. MTS nods continuously in line 11 while looking at her score, again signaling alignment.

Examples 1 and 2 have shown that evaluative stance-taking in instruction can be layered in different ways. In Example 1, the conductor performs a mocking enactment of the musicians'

```
09   C4     %speel is gewoon%     +& (0.4)    %di: %#da:+% dadida=
             just play
     c4                          +leans forward,
                                   then lengthens torso+
     c4     %..............%BH in- and upward%draws BH away from body%
     c4                           &raises eyebrows------------->
     fig                                                  #3
10   C4     =met ne schone aanzet,
             with a nice start of the note
11   C4     alles wat ge wilt moet navenant zijn,=
            everything you want needs to be in proportion
12   C4     =maar-
             but
13   C4     (0.4)
14   C4     %het is zo%&*((suppressed #sounds 1.0))%&*
            it's like
     c4     %.........%claws hands ifo torso,
                        pulls up shoulders--------%
     c4                &presses and puffs lips,
                         frown-raises eyebrows-----&
     c4                 *tilts head right and down---*
     fig                              #4
15   C4     %&*en het # KOmt niet.%&*
               and it doesn't come
     c4     %opens and pulls apart
              hands, drops shoulders%
     c4      &neutral face---------&
     c4       *center position------*
     fig              #5
```

Image 3 Image 4 Image 5 Image 6

```
16   C4     (0.5)
17   C4     ja?
            yes
18   C4     (0.9)
19   C4     probeer dat is +%  (0.5) %%#°hh+%
            try that
     c4                    +leans forward,
                             then lengthens torso+
                         %hands towards %hands away
                                 body%    from body%
     fig                                  #6
20   C4     drie °hh
            three
21   C4     ♫hh°
```

FIGURE 3
Example 1.

performance which leads to evaluations on two layers: one of the actual previous performance on the base layer and one of the image of a chicken laying an egg and connected associations on a pretense layer. The overarching stance object, which is the musicians' performance, overlaps with the target of the mockery, namely, the musicians. In Example 2, a different kind of stance-stacking is achieved through self-deprecation by the coach. Embedded into the hedged negative evaluation of the previous performance, there is a negative, mocking stance toward SG's own instructions on another layer. The examples illustrate the constant dynamic evolution of stance objects, which Iwasaki (2022, pp. 7–10) has highlighted previously in terms of their temporal unfolding. In addition to this

01 SG äh-
erm
02 SG (0.6)
03 SG ~$die zäsur# in dem piano die is super-=
the cesura in the piano that is great
sg ~smiles----------->
sg $tilts torso right and forward------->
fig #7

Image 7: SG

04 SG *=die is *$MIR jetz wo ich doch*$*so;*
that is now for me while I so much
sg *........*touches torso with LH* *...*
sg $nods repeatedly----->$
05 SG [*$((laughs 1.0))$&+%für zäsu#ren* werbe,*
advertise cesuras
sg *BH as if holding up a sign-----*,,,,,,,*
sg $sits up and
leans back$
mts &smiles----------->
am %smiles----------->
oh +smiles----------->
fig #8
06 MTS [mhm,
mhm

Image 8 (L-R): MTS, AM, AC, OH and SG

FIGURE 4
Example 2.

sequential development, we can see that mocking enactments coincide with a layering of objects and their evaluation at one moment in time, through the combination of serious and non-serious layers.

While in Examples 1 and 2, there is mostly congruency between who/what the stance object is in relation to the target of the mockery, the mocking enactments serve very different functions in terms of positioning of the subject who performs them. This intricate positioning process will be the focus of the next section.

4.2 Positioning

In the second part of the analysis, we turn to spontaneous conversations between friends and zoom in on the process of positioning. We scrutinize how the target of the mockery is positioned, as well as how the participants distance themselves from the stance object and the target of the mockery. Using two examples, we show that, similar to evaluation, positioning is a layered endeavor. In both Example 3 and 4, participants invoke other viewpoints instead of

```
07    SG    °h mir% is die §DA§ ein bissl& zu~ [viel,+
            for me it is a bit too much there
08    MTS                                      [bissl zu viel;=
                                                a bit too much
      sg    ------------------------------>~
      sg                  §nods§
      mts   --------------------------->&
      am    ----->%
      oh    ---------------------------------------->+

09    MTS   =m[hm?
              mhm
10    SG      [äh-
               erm
11    SG    weil §ICH§ fänd schön *wenns ein b-*
            because I would find it beautiful if it
      sg          §jerks          *raises hands
                 head back§           slightly*
12    SG    DA ein *BIsschen weiter schwingt.
            it keeps on swinging a bit
      sg          *LH PU, moves RH in arching movement up and out*
13    SG    &also es is d-  (.) *dass die we**lle weitergeht;
            so it's that the wave keeps going
      sg                  *small circle  *RH in arching
                                with RH*   movement up and out*
      mts   &nods repeatedly while looking at score---->>
```

FIGURE 5
Example 2.

(Example 3) or next to (Example 4) the viewpoint of the target of the mockery. In the first part of the analysis, we discuss who or what is enacted, what the stance object is, and who the target of the mocking is. Additionally, we analyze how participants position themselves.

Consider Example 3 below, taken from the coffee bar corpus (see Supplementary materials S3, S4). In this example, three friends from university (Michael, Noor, and Melanie) are discussing how much they like fresh soup but that it is so expensive to buy in the supermarket ready-made. First, we will examine lines 01–08, including the build-up toward the mocking enactment. Next, lines 09–13, including the mocking enactment, will be discussed.

In lines 01–03, Michael proposes that it is not that much trouble to make soup yourself if you buy an immersion blender. During his proposal, Noor bursts out in laughter (line 04), which is attended to by both Michael and Melanie. Melanie then explicates the laughable (lines 06–08) and mocks Michael's proposal, arguing that it is a ridiculous idea to spend so much money on "such a fucking thing." A discussion follows about the pros and cons of making your own soup (the omitted lines).

In line 04, when Noor bursts out in laughter (Figure 6 - Image 9), she produces the first negative evaluation, evaluating Michael's proposal to "buy your own blender" as laughable, and clearly distances herself from Michael's stance. At the same time, this opens the floor for Melanie to explicate the laughable and produce the first mocking utterance, which is not yet an enactment. Shifting pronouns from a generic "ge" (*you*) to the specific hypothetical of Melanie herself ("ik," *I*), she stages a layered evaluation of Michael's proposal. By pretending to go along with him and "just go and pay 70 euros for such a fucking thing," the contrast between two scenarios is invoked, in which paying 70 euros for a blender is an excellent idea versus a terrible idea, resulting in a stacked stance. The layered evaluation serves to position Michael's proposal as absurd. Furthermore, Melanie's use of the first-person pronoun "I," denoting her hypothetical self, in combination with her gaze averted to her hands, contributes even more to the distancing with respect to Michael's proposal (see also Haddington, 2006).

Subsequently, a serious discussion follows concerning the pros and cons of making your own soup with an immersion blender (the omitted seconds). In line 09, Michael argues in favor of making your own soup but is interrupted once more by Noor (line 10), who then produces the mocking enactment and is joined by Melanie (lines 11–13, images 10 and 11). Finally, in the silence that follows, Michael produces a compound shrug (line 14, Figure 7 - Image 12).

Continuing his argumentation (line 09), Michael resists Melanie's and Noor's line of thought (averting his gaze from Melanie) and positions his idea as obvious, as indexed by the palm open gesture with both hands (Marrese et al., 2021). In her interrupting turn then, Noor verbally sets up a stage for the hypothetical scenario and the enactment to come ("I can see that Melanie"). Note also the switch in pronouns similar to the first part, from the generic "ge" (*you*) in line 09 to "die Melanie" (*that Melanie*) in line 10. This switch is interesting in two ways. First, it presents a shift from a general proposal to a specific hypothetical scenario, emphasizing Melanie's role in it. Second, it functions to distance Melanie the addressee from "die Melanie" (*that Melanie*) involved in the fictional scenario that is about to be staged.

In the enactment, Noor uses her right hand to depict using an immersion blender, while visually attending to this action. At this point, it may be noted that the enactment itself is not staged by drawing on exaggeration (as is the case in Example 1 above). Rather, the depiction of this action serves to draw attention to the

```
01    Michael    op zich,
                 actually
02    Michael    als ge een mixer koopt,
                 if you buy a blender
03    Michael    is dat ook niet [veel moeite om & te ~maken hè,
                 it's not a lot of trouble to make it right
04    Noor                       [*((laughs loudly, 1.6 sec))#
      Noor                        *closes eyes, leans torso back, throws head back
                                   towards ceiling--->
      Melanie                                  &gazes at Noor--->
      Michael                                       ~gazes at Noor, smiles-->
      fig                                                  #9
```

Image 9 (L-R): Michael, Noor, Melanie

```
05                 (0.8)
06    Melanie      ik zal ~eens effe zeventig euro gaan [betalen=
                   I will just go and pay seventy euros
      Michael      ------>~gazes at Melanie--->
07    Noor                                              [<<laughing>ja>
                                                                 yeah
08    Melanie      =voor zo& een <<laughing> kut ding>&
                    for such a fucking thing
      Melanie              &gazes down at hands------>&

(5 seconds omitted)
```

FIGURE 6
Example 3.

incongruence between the act of blending soup and Melanie being the one performing the action (much like in Example 2 above). This incongruence gives rise to the positioning of the proposal as absurd, the positioning of Michael as the target by extension, as well as the distancing of Noor and Melanie from the stance object and target (i.e., the stacked stance).

Finally, following their enactments, Melanie and Noor cover their faces with their hands (Ford and Fox, 2010) visually distancing themselves from the scenario and dismissing the idea (Figure 7 - Image 12), while Michael continues to resist and produces a final shrug.

It is interesting to note the incongruence between the character of the enactment and the target of the mockery. Instead of depicting Michael, Noor positions Melanie as the character of the enactment. In isolation, the enactment in lines 10–12 could be interpreted as mocking Melanie. Melanie is positioned as a 'typical student' who does not put too much effort into cooking, whereas Michael positions himself as a responsible person who thinks ahead (e.g., about having food for 3 days, line 09). It is especially the misfit of the idea of Melanie blending soup, and the difference between Melanie and Michael, that gives rise to the mockery. However, in light of the preceding turns (lines 01–08), including both Noor and Melanie mocking Michael's proposal, we argue that the low likelihood of Melanie blending her own soup merely serves to highlight the absurdity of Michael's proposal, rather than staging Melanie as a target as well. This in turn highlights the notion that positioning with regard to the stance object and target is done locally in the interaction but also draws on and extends to positioning in a wider sociocultural field including membership categories (see Section 5.3).

Another tool that participants have at their disposal to distance themselves from the stance and target at hand, and which emerged prominently in our data set, is by including a viewpoint shift immediately following the mocking enactment. As an illustration, consider Example 4 (see Supplementary material S5). In this extract, three friends (Jilske, Yana, and Nikki) are talking about Lucas, a friend of theirs, who forgot to take his lab glasses to class and then came rushing into Jilske and Nikki's class to ask whether he could borrow one from them. Immediately preceding the excerpt, all participants are displaying what could be interpreted as *Schadenfreude* on the incident. In lines 01–03, Nikki adds to the anecdote that Lucas did bring his lab coat "in compensation". In lines 07–10 then, many things happen simultaneously. Nikki produces a mocking enactment, in which she depicts Lucas offering his lab coat "in compensation." Simultaneously, both Yana and Jilske evaluate the situation as a whole. Finally, in lines 14–15, Nikki adds another enactment, this time from her own viewpoint, mocking their friend once more.

Throughout the example, all participants position Lucas as the target of the mockery. In line 01, Nikki utters the statement that he brought his lab coat "ter compensatie" (*in compensation*), while producing a palm up open hand gesture (PUOH) as if presenting the lab coat, and a shoulder shrug, distancing herself from this action. While uttering "labojAs" (*lab coat*) with emphasis on "coat," she raises her shoulder again, highlighting the discrepancy between what was expected—to bring glasses—and reality—that he brought a coat. This

10.3389/fpsyg.2024.1379593

09 **Michael** **ja oké ~maar ge hebt wel~ eten voor ~(-) [drie da~gen hè,**
yeah okay but then you have food for three days right
Michael ------>~averts gaze from Melanie~ ~opens both palms~
10 **Noor** **[ik *zie die melanie=**
I can see that Melanie
Noor *moves head back--->
11 **Noor** **=$~daar al [%@staan *in &haar keu]ken#&**
standing there in her kitchen
Noor $smiles->>
Noor %lifts RH in fist, as if using an immersion blender-->
Noor ------------------>*tilts head, looks at hand--->
12 **Melanie** **[((laughs))** **]**
Melanie @smiles->>
Melanie &lifts RH in fist, as if holding a blender&
Michael ~presses and retracts lips--->
fig #10 & 11

Image 10: Noor Image 12: Melanie

13 **Noor** **<<laughing> &(zo)~staaf*mixen>%**
(like) immersion blending
Michael ----------------->~
Melanie &brings RH to head, palm on forehead-->>
Noor ---------------------->*shifts gaze to Michael, then closes eyes->>
Noor ------------------------------>%brings elbow to table and palm to forehead-->>
14 ~(1.3)#~
Michael ~presses lips forwards, shrugs left shoulder, both palms open~
fig #12

Image 12 (L-R): Michael, Noor, Melanie

FIGURE 7
Example 3.

is exploited further in line 02 when Nikki explains that "they did not need to bring that," a statement accompanied by another PUOH gesture and a head shake, presenting this as obvious and referencing the failed expectation. Both Yana and Jilske align with the stance at hand by laughing in lines 04–05, and Yana frown-raises and covers her mouth with her right hand (line 06, Figure 8 - Image 13). This gesture can be interpreted as displaying vicarious embarrassment (or *Fremdscham*) on Lucas' behalf, or self-censorship of a potentially inappropriate response to the event. In all, her response contributes to the evaluation of Lucas' actions as embarrassing and as such distancing herself.

In line 07, Nikki enacts Lucas ("*and he was like, at least something*"), which, by means of the raised eyebrows, PUOH gesture, and head shake and tilt (Figure 9 - Image 14), gives rise to a feeling of desperation, as if pleading for his case. The enactment highlights the inappropriateness of Lucas' action—the unsuitability of a lab coat to replace lab glasses—resulting in the mocking character. Simultaneously, Yana produces another enactment (line 09). In this enactment, it is unclear who the enacted character is—herself, Lucas, or a more general audience. What emerges can be described as a doing evaluating, in which not so much the positioning of a specific stance subject is relevant, but rather the evaluation of the situation as embarrassing, or a situation in which one would say "ouch." Nikki then adds another viewpoint (next to the viewpoint of the target of the mockery from line 07, and the viewpoint of a generic evaluator in line 09) in an enactment of her own (mocking) stance in response to the events. She positions herself as a calm, potentially authoritative figure, praising his effort ("goeie intenties," *good intentions*) while tilting her head, closing her eyes, and nodding (Figure 9 - Image 15). This layered evaluation presents a contrast with the first enactment of the desperate target of the mockery and even more directly distances the speaker from the target.

```
01    Nikki    ma' hij had (.) *&[in comp- ter] compensatie*& zijn *labojAs mee.*
               but he had in comp- in compensation his lab coat with him
02    Yana                          [wel schattig;]
                                     cute though
      Nikki                      *shrugs R shoulder----------*    *raises R shoulder*
      Nikki                       &gestures PUOH with RH------&
03    Nikki    *ma' die &mOEsten wij nie'& [meenemen.*
                but we didn't have to bring that
04    Jilske                               [((laughs))
      Nikki    *shakes head-------------------------*
      Nikki              &gestures PUOH with RH&
05    Yana     ~oh: $[((laughs))#
06    Nikki           [((laughs))
      Yana     ~frown-raises eyebrows----->
      Yana          $brings RH to mouth--->
      fig                          #13
```

Image 13: Yana

FIGURE 8
Example 4.

Jilske, finally, somewhat resists the mockery and affiliates with Lucas' position (in lines 08 and 13, saying "poor guy"). Indeed, after the extract, she continues with a defense of Lucas' actions, saying that she understands how you could lose your glasses. In Section 4.3 below, we will discuss how participants sometimes partially align with separate layers of the mocking stance.

In sum, this section described how participants (1) position the target of the mockery and (2) distance themselves from the stance object and target of the mockery. The layered evaluation (as presented in Section 4.1) can use absurd scenarios including analogy and imagery, as well as exploit the contrast between some expectations and reality, often drawing on broader membership categories and stereotypes. Together, this results in a layered positioning in which (a) the character of the enactment, (b) the stance object, (c) the target of the mockery, and (d) the interlocutors in the interaction are positioned. With regard to the role of different viewpoints in positioning, we saw that participants draw on layered viewpoints: Next to enacting the target of the mockery, participants can enact themselves to include their own viewpoint and position. In some cases, the target of the mockery is not the character of any enactment in the interaction at hand.

4.3 Alignment

In this section, we focus on the layering of alignment. As discussed in the previous examples, during mocking enactments, interactants construct a layering of both evaluations as well as positioning. Since multiple objects and multiple subjects are involved in the stance act, the negotiation of alignment can gain complexity. Not only can all present interactants (dis)align with each other and with the enacted character, but they can also do this for multiple evaluative layers toward the object separately (e.g., only align with the non-seriousness but not with the evaluation of the object).

In this section, we look at one longer sequence containing multiple mocking enactments that will illustrate the layering of alignment and the complexity of its negotiation.

In Example 5 (see Supplementary materials S6, S7), two signers (Susan and Donna) are discussing the ways people cheer for games of the Belgian national soccer team. When confronted with this topic in the elicitation task, Susan immediately states that she does not know anything about soccer and produces an away gesture, waving the topic away (Bressem and Müller, 2014), and expressing her stance on the topic both lexically and non-lexically. Donna, however, expresses her enthusiasm about the sport, smiling, and nodding while she states that she saw a game the day before. Subsequently, Susan states that this is not an interesting topic. The two immediately express disaligning stances on the topic in overlapping turns (lines 4–7), both epistemically and affectively. Then, Susan produces a mocking enactment: In line 08, she enacts herself looking at Belgian flags hanging out of the windows of apartment blocks and pointing at these flags (Figure 10 - Image 16). This enactment is followed by a second one, in which Susan makes a palm forward gesture, averts her gaze, turns her head to the right, and sticks her tongue out of her mouth, expressing her disgust. In this example, the first mocking enactment is viewpointed from the signer's perspective. The character of the enactment is the signer herself in the past, and the target, people who hang out Belgian flags during soccer games, is only implied. While enacting herself looking at apartment windows, feeling disgusted when she sees these flags hanging out of windows, she not only positions herself, evaluating this object. Rather, within the context of this interaction, she also positions people who hang a flag out of their window as the target of ridicule. As discussed in Section 4.2, there is an incongruence between the character of the enactment and the target of the mockery. During

10.3389/fpsyg.2024.1379593

07 **Nikki** **%<<laughing>dus~ hij [was zo &van]$,> [*(0.5) toch #ie:ts;%&***
so he was like *at least something*
Yana ---------->~
Yana ---------->$
Nikki %raises eyebrows---------->%
Nikki &gestures PUOH with RH------->&
Nikki *shakes and tilts back head*
fig #14

Image 14 (L-R): Jilske, Yana and Nikki

08 **Jilske** £[**ocharme;]** £
poor guy
Jilske £raises eyebrows£
09 **Yana** [**da's echt zo- oei; $~(0.3)$~**
that's really like ouch
Yana $draws in head$
Yana ~raises lower lip~
10 **Yana** **[((laughs))**
11 **Nikki** **[((laughs))**
12 **Jilske** £**ocharme**£
poor guy
Jilske £raises eyebrows£
13 **Nikki** ***ik dacht,**
I thought
Nikki *tilts head towards left--->
14 **Nikki** ***%#goeie% in[tenties*.**
good intentions
Nikki *nods---------*
Nikki %closes eyes%
fig #15

Image 15: Nikki

15 **Jilske** **[ik snap wel da' ge da' kwijt kunt een labobril.**
I get that you can lose those, lab glasses

FIGURE 9
Example 4.

these two mocking enactments, Donna smiles and, in overlap with the second enactment, points at herself. She signs that she herself hangs the Belgian flag out of her window during soccer games. In line 10, she continues by saying that she also decorates her car and that she likes doing that. As such, she makes clear that she is a member of the group of people that constitutes the target of the mockery in Susan's enactment. While she does that, Susan shakes her head and gasps, expressing her negative stance. Susan continues in line 11, saying that "it's horribly exaggerated." In this example, we can observe how the two signers do not align with each other, engaging in a heightened multimodal stanced discussion, combining a range of resources to express their disalignment in overlapping turns, in the run up to a mocking enactment. Moreover, it becomes clear that Donna is—possibly by accident—not only a subject who can align with the mocking stance but also the target of the mockery. Subsequently, the two elaborate on the topic, engaging in a 1-min discussion on the topic

10.3389/fpsyg.2024.1379593

01	**Susan**	**KNOW-NOTHING**
		I don't know anything about that
02	**Donna**	**SOCCER PS-1 STILL PS-1 FOLLOW PS-1**
		I still follow soccer games
03	**Susan**	**NOTHING NOTHING**
		Nothing, nothing
04	**Donna**	**CHEER YESTERDAY EVENING EXCITING [YESTERDAY EVENING] PS-1**
		Yesterday evening I cheered, it was exciting
05	**Susan**	**[INTERESTING NOT]**
		It's not interesting
06	**Donna**	**[UNTIL G:HE UNTIL C:12 HALF PS-1 PS-1**
		Until half past 12
07	**Susan**	**[APARTMENT PS-SWEEP HANG FLAGS APARTMENT**
		And there's flags hanging from the apartments
08	**Susan**	**PS-3 PS-3 PS-3 [pointing to various high locations, leaning forward, gazing up, as if pointing to flags hanging from apartment windows]**
		There, there, there
	fig	#16

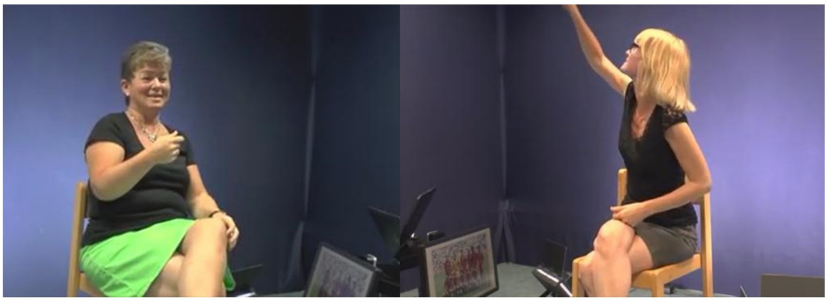


Image 16: (L-R): Donna & Susan

09	**Susan**	**[Palms forward, averted gaze, head turned to the right and tongue out of mouth, as if expressing disgust] PS-1**
10	**Donna**	**[PS-2 PS-1 G:HE HOME waving-flags ABOVE waving-flags CAR around-car**
		You, I have a flag at home, and around my car
11	**Susan**	**EXAGGERATE BAD**
		It's horribly exaggerated
12	**Donna**	**PS-1 IN CAR BALL RED BALL**
		I have a red ball in my car
13	**Donna**	BUT FUN PARTICIPATE PS-1
		But for me it's fun to participate

FIGURE 10
Example 5.

of using flags and decorations to cheer during soccer games, overtly expressing their strong disalignment (lines omitted in the transcript).

After engaging in a discussion, Susan signs that some people take a day off to go out and party the day after a soccer game, frowning, opening her eyes wide, and moving her body forward. Donna leans back, smiles, and nods, agreeing that this is true. With this enactment, Susan shifts her viewpoint, enacting the target of the mockery, distancing herself from the stance and target, as discussed in Section 4.2. Then, Susan produces another mocking enactment in line 16, enacting people who go out to party in the Belgian city of Aalst and, waving her hands from left to right, sticking her tongue out of her mouth, frowning, and leaning forward, as if shouting and partying (Figure 11 - Image 17). Donna aligns, smiling and saying that is true. Susan elaborates on this, clarifying that she is talking about young people who go out to party in the city center. She continues this enactment, showing how these people say that they do not want to work the next day and take a day off. With this second mocking enactment sequence, the target of the mockery narrows down to a more specific group of people. With the mocking enactments in lines 17 and 22, Susan evaluates the object of people who go out to party in Aalst and do not work in a layered manner, creating a stacked stance expression: The enactment serves as an evidential, presenting what she knows about these people, while negatively evaluating them through enacting their behavior in a non-serious, exaggerated manner. As such, in contrast to the previous enactment, she reduces the scope of the group that constitutes the target of the mockery, excluding Donna. Consequently, this mocking enactment sequence serves as a mitigation strategy with regard to the previous mocking enactment. While during the first enactment, the behavior of Donna was negatively evaluated, during this enactment, there is room for her to position

14	**Susan**	**ALSO MANY PS-3 TAKE TODAY [SPECIAL DAY^OFF**
		Many people also take a day off today, especially for that
15	**Donna**	**[SPECIAL DAY^OFF TAKE**
		take a day off especially for that
16	**Susan**	**BECAUSE KNOW YESTERDAY SOCCER FINISH**
		Because they know yesterday there was a soccer game
17	**Susan**	**[hands with limp wrist waving from left to right, tongue out of mouth, frown and body leaning forward,as if shouting and partying]**
	Fig	#17

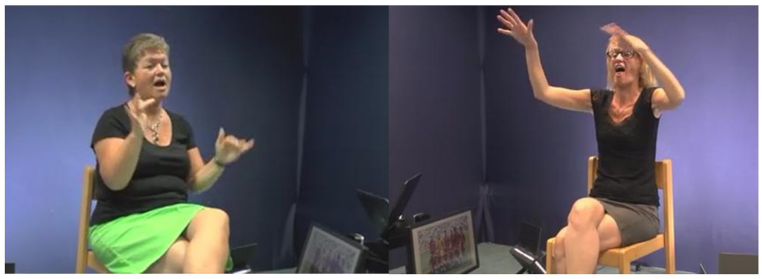


Image 17: (L-R): Donna & Susan

18	**Donna**	**[RIGHT PARTY AALST] PS-3**
		That's right, they party in Aalst
19	**Susan**	**[THEN PS-3**
		Then they
20	**Donna**	**[RIGHT AALST PARTY YES]**
		Right, they party in Aalst, yes
21	**Susan**	**[AALST PARTY YOUNG CENTER PS-3 PARTY PLACE FINISH BACK**
		In Aalst young people go to the center to party there and then they return
22	**Susan**	**WANT-NOT TODAY WORK NOT [uttered leaning backwards, enacting soccer fans]**
		I don't want to work today
23	**Donna**	**DAY^OFF TAKE PS-3 SOCCER**
		Taking a day off because of the soccer game
24	**Susan**	**[away gesture, face moved to the right, averted gaze, then crossing arms]**
25	**Donna**	**G:PU HAVE MANY PEOPLE PS-3**
		There's many people like that

FIGURE 11
Example 5.

herself toward an external target, allowing her to align with her interactant. Thus, the mocking enactment in line 17 constitutes a part of the negotiation of alignment on the topic as a whole and specifically with regard to the previous mocking enactment where one of the interlocutors was a member of the target group. However, in line 20, Donna again expresses that she knows people do take days off after going out to party, but she does not express her alignment on the mocking stance. While Donna aligns with Susan on both the epistemic layer and the non-serious layer of the mocking enactment with smiles, nods, and a lexical expression recognizing that the statement in itself is true, they do not reach alignment on the evaluation. As such, they only partially align on this stacked stance.

It appears, therefore, that the layering of the stance acts has implications for the expression of alignment. When interactants layer the evaluation and positioning in the stance act, they can also only partially align with these complex stances. If so, interactants seem to be more likely to elaborate on these stance acts. This layering of alignment relates to Clark's (1996) distinction between imagination and appreciation: "When there are two layers, the primary participants are to imagine the actions in layer 2, and appreciate the actions in layer 1" (Clark, 1996, p. 360). It is very well possible that interactants do imagine the enacted scene their interlocutor depicts but do not appreciate their actions on the base layer in the interaction. Indeed, our analysis confirms the findings of Holt (2016, p. 101) stating that layering of playful and serious stances presents the recipient with options in terms of responding to these playful first turns. They can align with playful aspects, with serious elements, or both.

When we consider these aspects of alignment with the presence of multiple subjects, layered stances, and possibly multiple objects, a complex negotiation may arise. As Iwasaki (2022) states, alignment or disalignment does not constitute a binary contrast but a fluid spectrum. The expression of a stance opens an opportunity for negotiation over the course of an interaction. However, we argue that on top of that, this spectrum of alignment and the possibility of partial alignment also exist in specific moments within the interaction.

Moreover, it becomes apparent how stance subjects, objects, and targets may overlap and be intertwined, and how this may evolve over the course of an interaction and implicate the negotiation of alignment. During the first mocking enactment, Susan supposed Donna to be a subject she could align with regarding the object of

"putting flags out of your window." However, it turned out that her interlocutor was not only another subject but also the object of the stance. The signers did not align, until Donna expressed her alignment by acknowledging the epistemic stance and the non-serious stance, but leaving her (negative) evaluation unexpressed.

Summing up, mocking enactments are complex multilayered stance acts that also entail a layering in expressing alignment, whereby interactants can align, disalign, or partially align. In the context of storytelling in interactions, it seems like alignment on every layer is desirable, if not necessary, for the progress of the interaction (Stivers, 2008; Peräkylä et al., 2015).

5 Summary and discussion

The aim of this study was threefold: First, we aimed to shed light on how enactments are used for mocking. Second, we asked which layers of stance-taking (on all components of the Stance Triangle) are at play. Third, we scrutinized how these enactments (and layers of stance) are multimodally constructed. In what follows, we will summarize and discuss our findings in light of the existing literature.

5.1 The use of enactment for mocking

Concerning the ways in which enactment can be used for mocking, a first finding relates to the creation of contrast. Our analysis shows that participants enact a target for the purpose of mockery (of either themselves or others) in an exaggerated or otherwise highly stylized way resulting in a contrast. This contrast could concern an incongruence between expectation and reality, or other norms to be adhered to, as is the case in Examples 1 and 5. In these examples, the interactants denounce specific characteristics of the target of the mockery by foregrounding and reshaping them in the way they represent these characters during enactments. However, as we saw in Examples 2 and 3, enactments can also be employed in a different way in mocking sequences. In these cases, the creation of the contrast does not lie in the design of the enactments themselves, in light of the expected or recognizable features of the target. Rather, these enactments serve to invoke a scenario that is mocked because of its absurdity.

A second finding concerns the use of viewpoint and character roles within the enactments. In prototypical cases of parodies to mock a target, the target is the character of the mockery, as in Examples 1, 2, 4, and 5. However, throughout the different data sets, we observed that signers/speakers also enact other characters that are not the target of the mockery, thus representing different perspectives, including their own viewpoint in response to the mocking enactment, or the viewpoint of other characters that are part of the depicted scenario (Examples 3 and 5). In other words, mocking enactments may not just consist of an enactment of the target but also of other characters who react to the target and stance object. This means that the target of the mockery, the stance object, and the enacted character do not necessarily overlap. In fact, a sequence with multiple enactments including viewpoint shifts, showing stances from different characters, can be employed to highlight contrasting stances.

5.2 The stance triangle in relation to stance-stacking

Through mocking enactments, stance objects are introduced on a non-serious level, outside of the shared base layer of the interaction, which results in a layering of evaluations and thereby a stacking of stances: In Examples 1 and 2, this was accomplished by introducing elaborate imagery embedded into a larger evaluation sequence, in Example 3 by invoking a (supposedly) absurd hypothetical scenario, and in Examples 4 and 5 by reporting on other characters' actions and stances.

Along with the layering of stance objects as well as evaluations, positioning can take different forms in mocking enactments. The question "who enacts whom in order to mock whom" is central to positioning through mocking enactment and links to what we summarized above about the use of different viewpoints in enactments for mocking. Variation occurs depending on whether the current signer/speaker positions themselves as the target of the mocking (Example 2), another co-present participant (Examples 1, 3, and 5), or an (imagined) external individual (Examples 4 and 5). Depending on the viewpoint and referent depicted in the enactment, this positioning can occur directly or by extension through the abovementioned layering of object and evaluation.

The layering of evaluations as well as positioning also has an effect on the negotiation of alignment. As discussed in the existing literature, the negotiation of alignment unfolds temporally across turns in interaction and should be regarded as a continuum rather than a dichotomy (Iwasaki, 2022). In the case of mocking enactments, an opportunity for (dis)alignment is opened up at one specific moment in time, where several layers are presented to the addressee to act upon. As mentioned in Sections 2.2 and 4.3, this links Clark's (1996, p. 360) distinction between imagination and appreciation. While the enacted scene can very well be imagined and aligned with on an epistemic or evidential level by the addressee of the mocking enactment, alignment with the mocking stance is not an obligatory consequence, which connects to the findings of Stivers (2008) on the difference between aligning and affiliating. In the case of layered stance acts, thus, participants can (dis)align, as well as partially (dis) align with the stance expressed (Holt, 2016).

Through the expression of stances on multiple interactional layers, interactants create stacked stances, with implications on every component of the Stance Triangle. Let us, therefore, zoom out to discuss what the model proposed by Du Bois has to offer for multimodal analyses of complex stance acts and longer stretches of interaction. In the existing literature on stance-taking, Du Bois' triangle has been much discussed and critically reflected on (Debras, 2015; Thompson, 2016; Iwasaki, 2022). Several authors have already drawn attention to the limitations of the model and suggested re-interpretations, such as Debras (2015) in the study on constructed dialog. Because of the additional subject that is introduced through enactment which lies outside of the interaction, she suggests to re-interpret the triangle as a tetrad. While this presents a relevant extension to the existing model for the phenomenon studied in said article, we suspect that for any given phenomenon and the changing complexity in terms of layering and inclusion of more or fewer subjects or objects, no geometrical form would suffice to account for the whole constellation and temporal unfolding.

Iwasaki (2022) highlighted the temporal transcendence of stance-taking in interaction and the constant evolution of all components of the Stance Triangle with every new stance act that is expressed. In this contribution, we add onto that by showing how during one stance act, several layers can be at play on the different components of the triangle, which can then either in part or as a whole be picked up in subsequent discourse. In that sense, we understand the Stance Triangle as a snapshot of a given stance act, or usage event, at one level of the interaction. Conceiving the Stance Triangle as a snapshot, a brief manifestation of complex processes that will already have evolved an instant later allows, as we have shown in our analysis, to tease apart participant roles, the parallel processes of evaluation, positioning, and alignment, and therefore, to distinguish layering on any of the vertices or sides of the triangle. Hence, in a mocking enactment, several triangles overlap, for the serious and the non-serious layer and possibly another for a reported stance within the enactment, resulting in stacked stances. We argue that these layers can only be teased apart fully with a multimodal approach, taking into account the full potential of interactional mechanisms at language users' disposal.

Rather than expanding the Stance Triangle, we, therefore, suggest taking it for what it is: a powerful model that breaks down complex cognitive, social, and interactive processes into a set of three components and processes, but which is inevitably limited and can never paint the full picture of stance-taking in complexly layered ways.

5.3 Multimodal design of mocking enactments

The third research aim concerned the multimodal design of mocking enactments. In this regard, we found that the enactments themselves were constructed multimodally, with a specific role for manual gestures in depicting the scene (e.g., clenched fists that open up in Example 1, or pointing to high locations to index flags in Example 5), in concert with many other resources (e.g., head tilts or shakes, eyebrow raises, and smiling) that have previously been described as having stance-related functions (see Section 2.1).

The multimodal design of individual enactments emerges locally in interaction. Therefore, it is difficult to identify specific patterns of multimodal resources that contribute to the layering of stances. Moreover, resources may serve multiple functions at the same time, which complicates assigning their use to the functions of, for example, either mocking or enacting (Cantarutti, 2021, p30–31). In Example 3, for instance, one of the interactants tilts her head during enactment, which can be an expression of her own stance on the enactment, a means to draw visual attention to the enactment, as well as a part of the enacted scene (see Figure 7 - Image 10). The use of multiple resources, however, constitutes a composite meaning [i.e., a multimodal "gestalt" (Mondada, 2014)], which results in a locally constructed stacked stance as a whole.

Furthermore, we found that mocking enactments are embedded in highly evaluative contexts that are constructed multimodally (e.g., gasping in Example 5, laughing in Examples 2 and 3, or covering the face with the hands in Example 3). As a result, a mocking enactment can often only be interpreted as a layered stance in relation to the preceding and succeeding interaction. We observed that in many cases, a negative evaluation has already been expressed or at least projected prior to the mocking enactment. In other words, one layer of the stacked stance may already be established before interactants add a second layer. This stacking can happen within the enactment but does not always have clear borders. In the context of the music instruction, for instance, the grammatical slot opener for the enactment space ("da's zo," *that's like*, Example 1) usually projects a negative evaluation which in that case is expressed as an enactment of exaggerated imagery.

Although the prototypical case of a mocking enactment consists of an exaggerated or otherwise "distorted" multimodal depiction of the target (as in Examples 1 and 5, cf. D'Errico and Poggi, 2016), we found that this is not always the case (as in Examples 2 and 4). Both these enactments were not performed using large gestures, a higher number, or more animated use of semiotic resources, and were not "distorted" in any other way. Instead, in these cases, the enactments merely seemed to allude to the hypothetical scenario that is mocked. We believe that this warrants further investigation. In what cases do participants choose to produce an enactment with more 'multimodal intensity', and in what cases is the enactment more minimal? A first step in this direction would be to quantify what precisely constitutes 'multimodal intensity' as there is currently no established operationalization of this phenomenon. Regarding the question of what influences the multimodal intensity of the mocking enactments, one possible answer would be the following. The easier it is to tease apart the non-serious layers (e.g., by the use of more or specific multimodal marking), the less aggressive a mocking enactment is for a recipient, and conversely, the more ambiguous the mockery is, the more negative response it may receive (Keltner et al., 2001; Yu, 2013). Research with an experimental setup could shed light on this topic.

Additionally, while enactment serves as a tool for local stance-taking in interaction, it simultaneously contributes to the construction of larger identity categories, including social and cultural identities, influencing individuals' perceptions of self and others both within and beyond immediate interactions. In these cases, an enacted individual can serve as a placeholder for a social group when interactants draw on stereotypes to mock these targets or their behavior. As in Example 5, a mocking enactment of an individual who goes out to party after a soccer game positions a whole social group as a target. The signer who produces this mocking enactment thereby excludes herself from this group and associated identities. While it is clear that mocking enactments contribute to the identity construction of self and others (Fischer and Simon, 2016; Gilbert, 2018; Van De Mieroop, 2020), both locally and beyond the borders of an interaction, the relation between these two dimensions remains largely unexplored. Cross-cultural and cross-linguistic research may uncover how mocking enactments are used to construct identities and (distancing) stances on a larger sociocultural scale.

6 Conclusion

The current contribution shows that mocking enactments offer fruitful grounds for the investigation of layered and stacked stances. We explored the use of enactment in mockery, finding that participants enact various characters, not only their targets but also incorporating other viewpoints. The shifts in perspective and highlighting of stacked and often contrasting stances show that the enacted character, the target of the mockery, and the stance object

may not necessarily overlap. While acknowledging the limitations of the Stance Triangle (Debras, 2015; Iwasaki, 2022), we view it as a useful framework to approach even intricately layered stance acts. Based on this model, we have shown that mocking enactments go together with layering on all components of the triangle, representing the base layer of interaction as well as potential non-serious and enacted layers. Furthermore, mocking enactments are embedded in highly evaluative contexts, which are indexed by a plethora of resources (bodily visual and/or vocal-aural, generic, or setting-specific). Finally, we found that although enactments *can* be staged in an exaggerated or highly stylized way, this is not necessarily always the case, an observation that warrants further (experimental) research.

Data availability statement

The datasets presented in this article are not readily available because the data analyzed in this study are subject to the following licenses/restrictions: The Corpus Flemish Sign Language is publicly accessible via www.corpusvgt.ugent.be. The other raw video data used in this study cannot be made available for privacy reasons. Anonymized fragments referred to in the analysis are available in the Supplementary material. Requests to access the datasets should be directed to FA, fien.andries@kuleuven.be.

Ethics statement

The studies involving humans were approved by Social and Societal Ethics Committee - KU Leuven. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

CdV: Writing – original draft, Writing – review & editing. FA: Writing – original draft, Writing – review & editing. KM: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The study reported in the present article was conducted as part of the MUST: Multimodal Stance Taking in Interaction project (C14/20/040) funded by KU Leuven.

Acknowledgments

The authors would like to extend their gratitude to Kristof De Weerdt and Carolien Doggen for helping with the inter-coder negotiation for the data in Flemish Sign Language (VGT) and to the reviewers for their comments on the first version of the manuscript. The members of the MUST team provided the authors with valuable feedback throughout this project. The authors also thank MuTh, the Javus Quartet (Marie-Therese Schwöllinger, Alexandra Moser, Anuschka Cidlinsky, and Oscar Hagen), and Clive Brown, Stefan Gottfried, and Johannes Meissl, for letting them study with the recorded sessions from MuTh Lab #8.

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/fpsyg.2024.1379593/full#supplementary-material

References

Ambadar, Z., Cohn, J. F., and Reed, L. I. (2009). All smiles are not created equal: morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *J. Nonverbal Behav.* 33, 17–34. doi: 10.1007/s10919-008-0059-5

Andries, F., Brône, G., and Vermeerbergen, M. (2022). Stance in Flemish sign language: a multimodal and polysemiotic phenomenon. *Belgian J. Linguist.* 36, 16–45. doi: 10.1075/bjl.00070.and

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). Multimodal stance-taking in interaction—a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Besnier, N. (1993). "Reported speech and affect on Nukulaelae atoll" in *Responsibility and Evidence in Oral Discourse*, 161–181.

Beukeleers, I., and Vermeerbergen, M. (2022). Show me what You've B/seen: a brief history of depiction. *Front. Psychol.* 13:808814. doi: 10.3389/fpsyg.2022.808814

Biber, D., and Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Interdiscip. J. Study Discourse* 9, 93–124. doi: 10.1515/text.1.1989.9.1.93

Boxer, D., and Cortés-Conde, F. (1997). From bonding to biting: conversational joking and identity display. *J. Pragmat.* 27, 275–294. doi: 10.1016/S0378-2166(96)00031-8

Bressem, J., and Müller, C. (2014). "The family of away gestures: negation, refusal, and negative assessment" in *Body—Language—Communication*. eds. C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin, Boston: De Gruyter Mouton), 1592–1604.

Brône, G., and Oben, B. (2015). InSight interaction: a multimodal and multifocal dialogue corpus. *Lang. Resour. Eval.* 49, 195–214. doi: 10.1007/s10579-014-9283-2

Burgers, C. (2010). Verbal irony: Use and effects in written discourse, [Doctoral Dissertation], Radboud University.

Cantarutti, M. N. (2020). The multimodal and sequential Design of co-Animation as a practice for Association in English Interaction, [Doctoral Dissertation]. University of York.

Cantarutti, M. N. (2021). *Co-animation and the Multimodal Management of Contextualisation Problems when Jointly 'Doing Being' Others*. *Social Interaction*. Video-Based Studies of Human Sociality, 4. doi: 10.7146/si.v4i4.128166

Caucci, G. M., and Kreuz, R. J. (2012). Social and paralinguistic cues to sarcasm. *Humor* 25, 1–22. doi: 10.1515/humor-2012-0001

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026

Clark, H. H., and Gerrig, R. J. (1984). On the pretense theory of irony. *J. Exp. Psychol. Gen.* 113, 121–126. doi: 10.1037/0096-3445.113.1.121

Clift, R. (2016). Don't make me laugh: responsive laughter in (dis)affiliation. *J. Pragmat.* 100, 73–88. doi: 10.1016/j.pragma.2016.01.012

Cormier, K., Smith, S., and Zwets, M. (2013). Framing constructed action in British sign language narratives. *J. Pragmat.* 55, 119–139. doi: 10.1016/j.pragma.2013.06.002

Couper-Kuhlen, E. (2012). "Exploring affiliation in the reception of conversational complaint stories" in *Emotion in interaction*. eds. A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 113–146.

D'Errico, F., and Poggi, I. (2016). "The bitter laughter". When parody is a moral and affective priming in political persuasion. *Front. Psychol.* 7:1144. doi: 10.3389/fpsyg.2016.01144

Dancygier, B. (2012). "Negation, stance verbs, and intersubjectivity" in *Viewpoint in language: A multimodal perspective*. eds. B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 69–93.

de Vries, C., Oben, B., and Brône, G. (n.d.). The coffee bar corpus: spontaneous triadic interactions between friends, Unpublished dataset.

Debras, C. (2015). "Stance-taking functions of multimodal constructed dialogue during spoken interaction" in *Gesture and speech in interaction—4th edition (GESPIN 4)*. eds. G. Ferré and M. Tutton (HAL), 95–100. Available at: https://hal.parisnanterre.fr//hal-01640486

Debras, C. (2017). The shrug: forms and meanings of a compound enactment. *Gesture* 16, 1–34. doi: 10.1075/gest.16.1.01deb

Drew, P. (1987). Po-faced receipts of teases. *Linguistics* 25, 219–253. doi: 10.1515/ling.1987.25.1.219

Du Bois, J. W. (2007). "The stance triangle" in *Stancetaking in discourse: Subjectivity, evaluation, interaction*. ed. R. Englebretson (Amsterdam, Philadelphia: John Benjamins Publishing Company), 139–182.

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15, 223–238. doi: 10.1515/cogl.2004.009

Engberg-Pedersen, E. (1993). *Space in Danish sign language: The semantics and Morphosyntax of the use of space in a visual language*. Hamburg: Signum, 406.

Ferrara, L., and Hodge, G. (2018). Language as description, indication, and depiction. *Front. Psychol.* 9:716. doi: 10.3389/fpsyg.2018.00716

Ferrara, L., and Johnston, T. (2014). Elaborating Who's what: a study of constructed action and clause structure in Auslan (Australian sign language). *Aust. J. Linguist.* 34, 193–215. doi: 10.1080/07268602.2014.887405

Fischer, R., and Simon, K. (2016). Pejorative aspects attributed to hearing people in signed constructed dialogue. In *Linguistik Aktuell/Linguistics Today*, vol. 228, eds., R. Finkbeiner, J. Meibauer and H. Wiese (Amsterdam: John Benjamins Publishing Company), 1–30.

Ford, C. E., and Fox, B. A. (2010). "Multiple practices for constructing laughables" in *Prosody in interaction* (John Benjamins), 339–368. Available at: https://www-jbe-platform-com.kuleuven.e-bronnen.be/content/books/9789027288462-sidag.23.27for

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor. Symb.* 15, 5–27. doi: 10.1080/10926488.2000.9678862

Gilbert, K. E. (2018). Evaluating evaluation: interactional positioning and embodied narration in police interviews. *Narrat. Inq.* 28, 1–29. doi: 10.1075/ni.17005.gil

Gironzetti, E. (2022). *The multimodal performance of conversational humor*. Amsterdam: John Benjamins Publishing Company.

Glenn, P. (2003). *Laughter in interaction* Cambridge University Press Available at: https://www.cambridge.org/core/books/laughter-in-interaction/4629463A15293CFEBD21EE70AAC966F2.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience*. Harvard University Press.

González-Fuente, S., Escandell-Vidal, V., and Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. *J. Pragmat.* 90, 26–47. doi: 10.1016/j.pragma.2015.10.002

Günthner, S. (1997). "Direkte und indirekte Rede in Alltagsgesprächen. Zur Interaktion von Syntax und Prosodie in der Redewiedergabe" in *Syntax des gesprochenen Deutsch*. ed. P. Schlobinski (Wiesbaden: VS Verlag für Sozialwissenschaften), 227–262.

Günthner, S. (1999). Polyphony and the 'layering of voices' in reported dialogues: an analysis of the use of prosodic devices in everyday reported speech. *J. Pragmat.* 31, 685–708. doi: 10.1016/S0378-2166(98)00093-9

Haddington, P. (2006). The organization of gaze and assessments as resources for stance taking. *Text Talk* 26, 281–328. doi: 10.1515/TEXT.2006.012

Hodge, G., and Ferrara, L. (2014). "Showing the story: enactment as performance in Auslan narratives" in *Selected papers from the 44th conference of the Australian linguistic society (2014)*, 372–397. Available at: https://www.semanticscholar.org/paper/Showing-the-story-%3A-Enactment-as-performance-in-Gawne-Vaughan/cc5fae968168cca1d33f146364cddf40dbbcdddd

Holt, E. (2013) in *There's many a true word said in jest': Seriousness and non-seriousness in interaction*. eds. P. Glenn and E. Holt (Bloomsbury), 69–89. Available at: http://www.bloomsbury.com/uk/studies-of-laughter-in-interaction-9781441164797/

Holt, E. (2016). Laughter at last: playfulness and laughter in interaction. *J. Pragmat.* 100, 89–102. doi: 10.1016/j.pragma.2016.04.012

Hsu, H.-C. (2021). When gesture "takes over": A multimodal analysis of speech-embedded nonverbal depictions in interaction [Doctoral Dissertation]. KU Leuven.

Hsu, H.-C., Brône, G., and Feyaerts, K. (2021). When gesture "takes over": speech-embedded nonverbal depictions in multimodal interaction. *Front. Psychol.* 11:552533. doi: 10.3389/fpsyg.2020.552533

Iwasaki, S. (2015). Collaboratively organized stancetaking in Japanese: sharing and negotiating stance within the turn constructional unit. *J. Pragmat.* 83, 104–119. doi: 10.1016/j.pragma.2015.04.007

Iwasaki, S. (2022). Stancetaking in motion: stance triangle and double dialogicality. *Text Talk* 43, 575–598. doi: 10.1515/text-2020-0222

Janzen, T. (2022). Embodied cognition: ASL signers' and English speakers' use of viewpointed space. *Lang. Contrast* 22, 227–258. doi: 10.1075/lic.00020.jan

Jefferson, G., Sacks, H., and Schegloff, E. (1987). "Notes on laughter in the pursuit of intimacy" in *Talk and social organisation*. eds. G. Button and J. Lee (Clevedon, England: Multilingual Matters), 152–205.

Jehoul, A. (2019). Filled pauses from a multimodal perspective. On the interplay of speech and eye gaze. [doctoral dissertation, KU Leuven]. Available at: http://limo.libis.be/primo-explore/fulldisplay/LIRIAS2814932/Lirias.

Kärkkäinen, E. (2012). On digressing with a stance and not seeking a recipient response. *Text Talk* 32, 477–502. doi: 10.1515/text-2012-0023

Keltner, D., Capps, L., Kring, A. M., Young, R. C., and Heerey, E. A. (2001). Just teasing: a conceptual analysis and empirical review. *Psychol. Bull.* 127, 229–248. doi: 10.1037/0033-2909.127.2.229

Mandel, D. (2022). Jemandem eine Stimme geben: Polyphone und polykinetische Äußerungen im Sprechen von Angesicht zu Angesicht [Dissertation, Uni Freiburg]. doi: 10.6094/UNIFR/238217

Marrese, O. H., Raymond, C. W., Fox, B. A., Ford, C. E., and Pielke, M. (2021). The grammar of obviousness: the palm-up gesture in argument sequences. *Front. Commun.* 6:663067. doi: 10.3389/fcomm.2021.663067

Maury-Rouan, C. (2011). ""Voices" and bodies: investigating nonverbal parameters of the participation framework" in *Integrating gestures: The interdisciplinary nature of gesture*. eds. G. Stam and M. Ishino, vol. *4* (John Benjamins Publishing Company), 309–320.

Meissl, K., Sambre, P., and Feyaerts, K. (2022). Contrast pairs in orchestra instruction as a window into the multimodal expression of stance. *Papers of the Linguistic Society of Belgium* 16. doi: 10.61430/MAAG6888

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *J. Pragmat.* 65, 137–156. doi: 10.1016/j.pragma.2014.04.004

Mondada, L. (2018). Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51, 85–106. doi: 10.1080/08351813.2018.1413878

MuTh. (2021). MuTh lab #8: Beethoven im Brennglas.

Niemelä, M. (2010). The reporting space in conversational storytelling: orchestrating all semiotic channels for taking a stance. *J. Pragmat.* 42, 3258–3270. doi: 10.1016/j.pragma.2010.06.015

Norrick, N. R. (2004). Hyperbole, extreme case formulation. *J. Pragmat.* 36, 1727–1739. doi: 10.1016/j.pragma.2004.06.006

Nota, N., Trujillo, J. P., and Holler, J. (2021). Facial signals and social actions in multimodal face-to-face interaction. *Brain Sci.* 11:1017. doi: 10.3390/brainsci11081017

Peräkylä, A., Henttonen, P., Voutilainen, L., Kahri, M., Stevanovic, M., Sams, M., et al. (2015). Sharing the emotional load: recipient affiliation calms down the storyteller. *Soc. Psychol. Q.* 78, 301–323. doi: 10.1177/0190272515611054

Pfänder, S., and Couper-Kuhlen, E. (2019). Turn-sharing revisited: an exploration of simultaneous speech in interactions between couples. *J. Pragmat.* 147, 22–48. doi: 10.1016/j.pragma.2019.05.010

Pfeiffer, M., and Weiss, C. (2022). Reenactments during tellings: using gaze for initiating reenactments, switching roles and representing events. *J. Pragmat.* 189, 92–113. doi: 10.1016/j.pragma.2021.11.017

Pillet-Shore, D. (2020). When to make the sensory social: registering in face-to-face openings. *Symb. Interact.* 44, 10–39. doi: 10.1002/symb.481

Poggi, I. (2015). The Conductor's parodies. A multimodal strategy for error management in music rehearsals. *J. Multimodal Commun. Stud.* II, 9–21.

Priego-Valverde, B., and Rauzy, S. (2023). "Facial gestures and laughter as a resource for negotiating humor in conversation" in *Interactional humor: Multimodal design and negotiation*. ed. B. Priego-Valverde (Berlin, Boston: De Gruyter Mouton), 131–168.

Schrooten, S., and Feyaerts, K. (2020). *Conducting fanfare orchestras. A multimodal corpus*. Leuven: KU Leuven MIDI.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., et al. (2009). Gesprächsanalytisches Transkriptionssystem (GAT 2). *Gesprächsforsch* 10, 152–183.

Shaffer, B. (2012). "Reported speech as an evidentiality strategy in American sign language" in *Viewpoint in language*. eds. B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 139–155.

Soulaimani, D. (2018). Talk, voice and gestures in reported speech: toward an integrated approach. *Discourse Stud.* 20, 361–376. doi: 10.1177/1461445618754419

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stoeckl, H., and Messner, M. (2021). Tam pam pam pam and mi – fa – sol: constituting musical instructions through multimodal interaction in orchestra rehearsals. *Multimodal Commun.* 10, 193–209. doi: 10.1515/mc-2021-0003

Tannen, D. (1986). "Introducing constructed dialogue in Greek and American conversational and literary narrative" in *Direct and indirect speech*. ed. F. Coulmas, vol. *31* (Berlin, New York: De Gruyter Mouton), 311–360.

Thompson, G. A. (2016). Temporality, stance ownership, and the constitution of subjectivity. *Lang. Commun.* 46, 30–41. doi: 10.1016/j.langcom.2015.10.010

Thompson, S. A., and Suzuki, R. (2014). Reenactments in conversation: gaze and recipiency. *Discourse Stud.* 16, 816–846. doi: 10.1177/1461445614546259

Van De Mieroop, D. (2020). A deontic perspective on the collaborative, multimodal accomplishment of leadership. *Leadership* 16, 592–619. doi: 10.1177/1742715019893824

Van Herreweghe, M., Vermeerbergen, M., Demey, E., De Durpel, H., Nyffels, H., and Verstraete, S. (2015). Het Corpus VGT. Een digitaal open access corpus van videos and annotaties van Vlaamse Gebarentaal, ontwikkeld aan de Universiteit Gent ism KU Leuven. Available at: http://www.corpusvgt.be; http://hdl.handle.net/1854/LU-6973686.

Vandenitte, S. (2022). Making referents seen and heard across signed and spoken languages: documenting and interpreting cross-modal differences in the use of enactment. *Front. Psychol.* 13:784339. doi: 10.3389/fpsyg.2022.784339

Wang, Z., Fan, W., and Fang, A. C. (2022). Lexical input in the grammatical expression of stance: a Collexeme analysis of the INTRODUCTORY IT PATTERN. *Front. Psychol.* 12:762000. doi: 10.3389/fpsyg.2021.762000

Weeks, P. (1996). A rehearsal of a Beethoven passage: an analysis of correction talk. *Res. Lang. Soc. Interact.* 29, 247–290. doi: 10.1207/s15327973rlsi2903_3

Yu, C. (2013). Two interactional functions of self-mockery in everyday English conversations: a multimodal analysis. *J. Pragmat.* 50, 1–22. doi: 10.1016/j.pragma.2013.01.006

Zlatev, J. (2019). "Mimesis theory, learning, and Polysemiotic communication" in *Encyclopedia of educational philosophy and theory*. ed. M. A. Peters (Singapore: Springer Singapore), 1–6.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 05 July 2024
DOI 10.3389/fpsyg.2024.1429232

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Dusan Stamenkovic,
Södertörn University, Sweden
Sümeyra Tosun,
Medgar Evers College, United States

*CORRESPONDENCE
Alan Cienki
✉ a.cienki@vu.nl

RECEIVED 07 May 2024
ACCEPTED 26 June 2024
PUBLISHED 05 July 2024

CITATION
Cienki A (2024) Variable embodiment of stance-taking and footing in simultaneous interpreting.
Front. Psychol. 15:1429232.
doi: 10.3389/fpsyg.2024.1429232

COPYRIGHT
© 2024 Cienki. 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.

Variable embodiment of stance-taking and footing in simultaneous interpreting

Alan Cienki*

Department of Language, Literature, and Communication, Faculty of Humanities, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Previous research has argued that consecutive interpreters constitute laminated speakers in the sense that they engage with different kinds of footing at once, representing another's point of view through their words in another language. These multiple roles also play out in their gesturing, as they sometimes indicate deictically who is the source of the ideas and stances they are expressing (the principal). Simultaneous interpreters, though, often work in an interpreting booth; they are often not seen by the audience, yet many of them gesture, sometimes frequently. How are simultaneous interpreters using gesture in relation to stance-taking and footing? We consider the case of simultaneous interpreters rendering popular science lectures between (both to and from) Russian (their L1) and either English or German (their L2). Though only hearing the audio of the lectures, the interpreters produced many gestures, which were analyzed for their function. Some representational and deictic gestures appeared to clearly involve the interpreter as the principal (writing numbers with one's finger to help remember them or pointing to two places on the desk to keep track of two different quantities mentioned). Other representational and deictic gestures are ambiguous as to whether they are enacting what the interpreter may have imagined what the lecturer did or whether they arose out of the interpreter's own thinking for speaking (e.g., tracing the form of a bird being mentioned or pointing to an empty space when the lecturer was referring to a graph). Pragmatic gestures, showing one's stance toward the topic of the talk, were the most ambiguous as to the footing, reflecting how the interpreter may be engaged in fictive interaction with their imagined audience. Self-adapters, however, more clearly involve the interpreter as the principal, as such actions are known to support cognitive focussing and self-soothing. In sum, we see varying degrees of clarity as to whose stance and principal footing simultaneous interpreters are expressing bodily as laminated speakers. The variable ambiguity can be attributed to the nature of gesture as a semiotic system, the functions of which are more often dependent on co-occurring speech than vice versa.

KEYWORDS

stance, footing, gesture, simultaneous interpreter, pragmatic, self-adapter

1 Introduction and theoretical background

1.1 Interpreting, gesturing, and thinking for speaking

If we reflect on the professional activity of simultaneous interpreting between two spoken languages, it is actually a very unusual activity. Interpreters are speaking while listening, and their speaking is in a different language than that being listened to. What the interpreters are speaking about does not concern their own ideas, opinions, or feelings, but those of someone else. Furthermore, simultaneous interpreting of spoken languages as it is usually performed in a conference setting or during a lecture often involves the interpreter being located out of view of the people being communicated to, i.e., those who are hearing the interpreter. Simultaneous interpreters are often seated in a booth, usually in the back of the room, and speak into a microphone so that they can be heard by those in the room equipped with headsets used for this purpose.

Given these unusual factors of this form of communication, it might be surprising to learn that many interpreters are gesturing with their hands while interpreting, even though there is no one viewing them (Stachowiak-Szymczak, 2019; Cienki and Iriskhanova, 2020; Martín de León and Fernández Santana, 2021). It is not unusual for people to gesture while speaking even when they cannot be seen by their interlocutor, as we know from the phenomenon of people gesturing while talking on the telephone. The gesturing in such cases has been argued to be an inherent part of the role of visual demonstration in dialogue, even when one cannot be seen by one's interlocutor (Bavelas et al., 2008). The gesturing can also be seen as tied to processes of conceptualization (Kita et al., 2017), which includes processes of thinking for speaking. Specifically, Slobin (1987, 1996) claims that there is a special form of thought, thinking-for-speaking, that is mobilized when expressing one's ideas verbally. It needs to adapt to the lexical and grammatical options that are available in the language being used in the moment. McNeill (1992 and elsewhere) calls the smallest unit of thought that has the capacity to grow into an utterance the "growth point." When engaged in spontaneous talk, growth points successively develop and unfurl into speech and sometimes also into gesture in a dialectic process, whereby the lexical and grammatical forms afforded by the given language and the cultural constraints on the gestures influence each other in the microgenetic processes of their expression (McNeill, 1992, ch. 8). "Gesture contributes material carriers to thinking-for-speaking," as McNeill and Duncan (2000, p. 157) argue.

Note, though, that the context in which simultaneous interpreters work, rendering others' ideas in another language in a monologic fashion, is quite different from that of dialogic interaction. Furthermore, the kind of thinking involved in simultaneous interpreting is quite different from that involved in developing and expressing one's own ideas. Not only is their speech not about their own ideas, but simultaneous interpreters' later recall of what they heard while interpreting is significantly worse than it is when listening and not interpreting (Darò and Fabbro, 1994). This is in line with the view that interpreting need not involve (and most likely often does not involve) full semantic processing as much as a shallower form of lexical access (Gernsbacher and Schlesinger, 1997). Simultaneous interpreting is also an unusual form of communication in that the interpreters are not to express their own stance toward what they are saying (Setton and Dawrant, 2016, p. 344). This is something explicitly taught in interpreter training; for example, the "Practical Guide for Professional Conference Interpreters" of the International Association of Conference Interpreters (AIIC) states, "the interpreter must never betray any personal reaction to the speech, be it skepticism, disagreement, or just boredom." Instead, the interpreter's loyalty is owed to the speaker "and to the communicative intent that the speaker wishes to realize, whatever the speaker's position or point of view" (AIIC, 2004/2012).

1.2 Stance-taking and footing

Stance-taking (Biber and Finegan, 1988, 1989) involves different aspects of the speaker's attitudes toward their message, e.g., the degree of certainty about what one is communicating (epistemic stance), the importance of the information and the degree to which it is in focus (relevance stance), or one's affectual relation to what is being communicated. In interaction, people can express their stance not only verbally, but also through visible behavior of different kinds, e.g., through their bodily posture (e.g., pulling one's torso back when disagreeing with an interlocutor), as a form of physical stance-taking; facial expressions (e.g., raised eyebrows questioning another's claims); and manual gestures. The role of the latter has been explored in a number of studies, from Kendon's (1995, 2004, 2017) consideration of the functions of pragmatic gestures to the work on what Bressem and Müller (2014) and colleagues (e.g., Ladewig, 2014) have called "recurrent gestures." Among the German speakers focused on in the latter studies, these include, for example, a palm-up open hand with a clockwise rotation to indicate vagueness or uncertainty, a brushing away movement for negative assessment, and a stretched index finger held upward to mark attention to what is being said (Bressem and Müller, 2014). Many of the families of gestures (Kendon, 2004) that these examples belong to are also recognized in other European cultures [e.g., French (Calbris, 1990), Italian (Poggi, 2014), and Russian (Grishina, 2017)], and we are just beginning to learn about other types of recurrent gestures in non-European cultures, e.g., Chinese (Harrison, 2021), Hausa (Will, 2022).

In addition to stance, there is the role that one has in interaction, which encompasses stance-taking. This is what Goffman (1981) characterized as footing. Goffman (1981, p. 128) notes that with footing, a "Participant's alignment, or set, or stance, or posture, or projected self is somehow at issue." If we take the case of spoken interaction, the footing most immediately related to the usage event (Langacker, 1988) of speaking is that of the *animator*, the role of being the person talking. Goffman (1981, p. 144) puts it plainly as the role of being "the sounding box in use" from which the utterances come. In some cases, though, the words being uttered might have been pre-determined and selected by someone else, as when a politician reads a speech written by a speechwriter. Goffman (1981) calls this role that of the *author*, the agent "who has selected the sentiments that are being expressed and the words in which they are encoded" (*ibid.*). Beyond that, one or more people might be responsible for the content of the words being uttered in terms of having epistemic authority over them. Goffman calls this footing that of the *principal*, the party "whose position is established by the words that are spoken, someone whose beliefs have been told" (*ibid.*). We can say that the principal is the one whose stance is expressed. Returning to the example of a politician delivering a speech, the party that Goffman refers to might be the political party that the politician represents.

1.3 Interpreters as laminated speakers

The various kinds of footing are not mutually exclusive. A leader of a political party who writes and delivers his or her own speech embodies all three roles. In other contexts, we may variably be taking on one or more forms of footing. Goffman (1981) observes that conversation (p. 154) and, more generally, experience itself (p. 156) are, as he calls it, laminated at various times, involving more than one kind of footing. Goodwin and Goodwin (2004) extend this to say that what Goffman presents is an "analytically powerful model of a laminated speaker" (p. 223) in terms of the different kinds of possible footing one may inhabit.

Considering the case of consecutive interpreters (dialogue interpreting), i.e., who render speakers' utterances after they have produced them, Vranjes and Brône (2021) point out that such interpreters are laminated speakers of a special sort. The interpreter is the animator of the words they are speaking, but the person whose words they are interpreting is the principal. However, who is the "author" of the interpreter's words? Vranjes and Brône argue that the author is both the interpreter and the one whose utterances are being interpreted, and in this sense, the interpreter is a laminated speaker. But consecutive interpreters negotiate the interaction not only verbally between the people who speak two different languages but also co-verbally, using eye gaze direction, head movements, and deictic hand gestures. They note (Vranjes and Brône, 2021, p. 97), "Our analysis reveals that interpreters have a repertoire of multimodal resources at their disposal to layer their utterances and draw attention to the principal while rendering the talk."

1.4 The research questions of the present study

In simultaneous interpreting, though, as discussed above, the interpreter is normally not physically in the interactional space with the speaker of the source text and the audience hearing the interpreter's renderings in the target language, and in fact is usually not even visible. Yet, such interpreters are gesturing in many cases. This gives rise to several questions concerning the relations between gesture, footing, and stance, which will be examined here, specifically:

1) How does the use of different functions of gesture during simultaneous interpreting relate to the role of the interpreter as a laminated speaker? In what ways is this similar to or different from the situation with consecutive interpreters, as discussed in Vranjes and Brône (2021)?
2) How are simultaneous interpreters using gesture in relation to stance-taking, as discussed above in section 2.1?

3a) Whose stance (which footing?) is being expressed in any given interpreter's gestures moment by moment in the process of rendering the original speaker's utterances?
3b) Can this even be determined in the context and conditions of simultaneous interpreting?

As we will see below, considering the different functions of gestures and viewing them through the lenses of footing and stance-taking can help us gain further insights into what may be involved in simultaneous interpreters' processes of thinking for speaking.

2 Method of data collection

Forty nine simultaneous interpreters were involved in the study conducted at a university in Moscow between 2019 and 2021. All were native speakers of Russian (Russian as L1), residing in Russia, mostly in Moscow. Twenty nine of them (13 female) (average age = 33 years old) were experienced interpreters working between Russian and English (average 9 years' experience in interpreting) and 20 (7 female) (average age = 33) between Russian and German (average 10 years' experience). The materials they were asked to interpret in each case were two ten-minute excerpts from science lectures originally delivered to audiences of laypeople: one in Russian which they were asked to interpret into their main second language or L2 (be that English or German) and one in their L2 which they interpreted into Russian. All of the lectures concerned issues around biodiversity on the planet and the extinction of species of animals. The Russian lecture in each case, from the popular science website PostNauka, addressed the question, "Is there a threat of a sixth mass extinction of species?"[1] The lecture in standard British English, a TEDx Talk, was on "Mass extinctions and the future of life on Earth."[2] The German lecture, from the ARD television's Mediathek website, was entitled "The end of evolution."[3]

The interpreters were provided with vocabulary lists several days in advance with discipline-specific terminology from the videos and suggested translations into Russian and the L2, as appropriate. However, they were not allowed to bring any materials (such as paper, pens, or mobile phones) with them into the interpreting booth in which they were recorded. The reason for this is that we[4] were interested in how they handled the cognitive load of the interpreting sessions unencumbered by external tools. This allowed for a uniform condition across participants (i.e., no variation in terms of what external resources they might use) which also afforded studying their free-handed gestural behavior.

Written informed consent was obtained from all the participants in advance, and all interpreters were assigned participant numbers to anonymize reference to them. They were not informed in advance that our study was focused on gesture use and initially were only told that we were interested in analyzing interpreting behavior. The sessions were conducted in an interpreting booth in an otherwise empty classroom used for training interpreters. The interpreter sat on a chair with no armrests at the small desk built into the interpreting booth. Three video cameras recorded each interpreting session. A Sony HRX-NX30P video camera with a Sony ECM-XM1 directed microphone attached was placed on a tripod to the side behind the interpreter, pointed downward to provide an over-the-shoulder view of the interpreter's hands on the desk. On the far edge of the desk, a

1 "Существует ли сегодня угроза шестого массового вымирания видов?" https://postnauka.ru/video/49851, lecturer: Nikolai Dronin.

2 https://www.ted.com/talks/michael_benton_mass_extinctions_and_the_future_of_life_on_earth?language=en, lecturer: Michael Benton.

3 "Das Ende der Evolution" https://www.ardmediathek.de/video/tele-akademie/prof-dr-matthias-glaubrecht-das-ende-der-evolution/swr/Y3JpZDovL3N3ci5kZS9hZXgvbzEyMDkzOTk/, lecturer: Matthias Glaubrecht.

4 The first-person plural pronoun 'we' will be used as the research was conducted in collaboration with members of the lab listed in the Acknowledgments.

small GoPro HERO3+ Silver camera was placed facing the interpreter. This recording angle gave a close-up view of the interpreter's hands and face. In addition, interpreters wore Tobii eye-tracking glasses while performing the task, which provided a view through the glasses of where interpreters were looking, but this viewpoint is not of concern for the present analysis.

Each participant interpreted a 10-min segment from the talk in Russian into their L2 and from the talk in their L2 into Russian. The order of the tasks was randomly varied per participant. Importantly, interpreters only heard the audio of the lectures, played to them via headphones attached to a laptop placed out of view on a small stool to the side of their interpreting desk. They did not see the video of the lectures (the video was not even played off of the laptop—just an audio file of the lecture was used). This was done so that the interpreters could not see the actions of the speaker and so would not be influenced by their gestures (they could not copy them). Each time, a one-minute warm-up portion of the audio that preceded the upcoming 10-min portion was played in order to allow the researchers to adjust the audio to the interpreter's wishes and to allow the interpreter to get used to the lecturer's voice and speaking rate. Then the researcher began the 10-min portion of the lecture, closed the door of the interpreting booth, and moved to a part of the classroom out of view of the interpreter (since the interpreter's desk faces a large glass window in the door, looking into the empty classroom).

After the interpreting sessions, participants were debriefed about the study and they were allowed to choose how we could use the recordings of their interpreting, with permission options ranging from the maximum (being allowed to post audio or video clips of their sessions on academic websites) to medium (permission to post or publish screenshots of them from the videos) to the minimum (permission to only publish drawings of their gestures).

3 Methods of analysis

The recordings of the two ten-minute sessions from each of the 49 interpreters results in 16 h and 20 min of data. For practical reasons, 2 min were taken from each of the 98 videos for detailed analysis—one near the beginning of the session, after the interpreter had gotten into the flow of the task (minute 3:00–3:59) and 1 min later in the session (minute 8:00–8:59). The videos were imported into the ELAN[5] software (Sloetjes and Wittenburg, 2008) for analysis. The speech was transcribed using standard orthography for the given language. Gesture units were annotated with each including any preparatory phase, stroke phase, and post-stroke hold, if there was one (following Kendon, 2004).

Gesture were coded for functions in the context of the interpreter's speech using a system adapted from those used in Müller (1998), Cienki (2010), and Bressem et al. (2013). Though gestures are often multifunctional in nature (viz. Kok et al., 2016), we aimed to identify the primary function of each gesture, as described in brief below. In cases where interpreters were gesturing with two hands at once, and the two hands were seen to be realizing different functions, we coded the function of the dominant hand, that being the hand with which the interpreter gestured the most in the recording. The resultant code book consisted of the following categories of gesture use: representational, deictic, pragmatic, and as an adapter.[6] The following descriptions are abbreviated versions of those from the code book.

- The representational function is accomplished through depiction of some content of the speech. This was assessed if any one of five modes of representation (adapted from Müller, 1998, 2014a) was used. These entail either *acting* as it might when performing an action involving an object (such as moving one's fingers up and down as if typing on a keyboard); moving one's open hands as if touching the surface of an object (so-called *molding*); keeping one or both of one's open hands in a position with palms facing each other or with the palm up as if *holding* an object; *tracing* a shape or line with one's fingertip(s); or using the hand to as if become an object, as when one's index and middle fingers are extended straight and *embody* a pair of scissors by separating and closing together again.
- The deictic function is accomplished through one or more extended fingers being used to point in a direction or to touch a surface (such as the interpreter's desk) to identify a spatial location, an imagined referent, or a moment in time.
- Pragmatic functions include various types. These include performative functions (showing whether one is posing a question, making a denial or an offer, etc.), parsing functions (e.g., indicating topicalization or commenting via one's utterance), and modal functions (including negation, intensification, evaluation, etc.) (Kendon, 2004, pp. 281–282). What Kendon calls modal functions involve showing one's attitude toward the current topic of the talk. Here we see the expression of epistemic stance, relevance stance, and affectual or attitudinal stance. Bressem and Müller (2014), for example, show how for German speakers a wavering open hand can express uncertainty or doubt (epistemic stance), beats can emphasize words being spoken (relevance stance), or an open hand, palm facing down or away from the speaker and moving laterally can express dismissal or rejection (attitudinal stance). Whereas the representational function relates directly to the semantics of the speech (occurring with or beginning just after the start of the gesture unit), and therefore such gestures may be unique in form, pragmatic gestures occur across many contexts, with similar groups of forms expressing related functions (thus the name "recurrent gestures"). Many gestures serve both representational and pragmatic functions, such as metaphorically holding a referent as if it were an object in the hands while also performatively offering the imagined referent to the addressee; therefore, in our study, the category of pragmatic function was reserved for cases when it was clear that the primary function was not that of representation.
- Adaptive functions, as discussed in Ekman and Friesen (1969) involve either self-adapters or other-adapters. Self-adapters are

5 From the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. https://archive.mpi.nl/tla/elan.

6 The American English spelling 'adapter' is used here, but note (if searching the literature) that much research on them employs the British English spelling 'adaptor.'

inwardly oriented movements, involving self-touching of some kind. Other-adapters are externally oriented and entail touching some object, such as rubbing the desk. Adapters may consist of discrete, one-off actions, such as quickly scratching oneself or pushing back one's hair, or sustained actions, such as rubbing one's fingers together for an extended time. Note that this category of adaptive functions is excluded from many gesture studies because of researchers' focus on referential and pragmatic functions of gesture. However, in the present study, we included them because of their prominent role, as discussed below.

The video analysis was performed by nine researchers in the project, who worked in three teams of three members each. The videos were first divided among the three teams for the gesture annotation and function coding. Consensus checks on the coding were conducted within each team. The ELAN files with the videos were then exchanged with other teams who checked if they agreed with what were annotated as gesture units and performed an independent coding of the functions of the annotated gestures. Discrepancies between annotations and coding were then discussed and resolved at regular meetings of the entire research group. The method here was inspired by that described in Stelma and Cameron (2007); in their case, it concerned the coding of intonation units in transcribed talk. The method they used, and that they recommend, involved annotation and re-annotation of a transcript by a given individual, with refinements over time based on consultation with other experienced researchers. In our case, however, the individual annotations were not created and coded for function by one individual, but rather were done independently by the three members of one team and three members of another team (thus six coders) to check if the consensus within each of the two teams matched. Final resolution of any remaining problem cases took place through discussion with all 10 project members. Any amendments to the code book resulting from clarifications coming out of the discussions of the cross-checking were then applied across all the videos. This procedure was followed for all the coding, not just with a small percentage of the data, as is often the case for a cross-check of inter-annotator agreement.

4 Results

As reported in more detail below, the vast majority of the 3,719 gestures produced over all of the interpreting sessions either had a primarily pragmatic function or (self-)adaptive function. Far fewer gestures served a representational or deictic function; below we will consider these two categories together as constituting different types of referential function. We will consider the possible reasons for this distribution of gesture functions in relation to how they relate to footing.

If we consider Goffman's account of footing in talk and translate it to gesture, we can say the following. The animator can be seen as the one moving who is producing the gestures. Therefore an interpreter gesturing while interpreting is the animator of the gestures they are producing. The principal can be said to be the party to whose position, stance, and/or beliefs the gestures attest. Therefore, the question to be discussed below is: whose stance do the interpreters' gestures reflect? Their own, or the imagined stance of the original lecturer being interpreted? Finally, there is the footing of the author. We could say that that is the agent who puts together, composes, or scripts the gestures that are produced. This would make sense in the context of a play or a movie, for example, where the director and/or the actor decide in advance what gestures will be produced when, or for a public speaker who has been advised by a communication consultant on how to gesture. However, in the context of simultaneous interpreters, this footing is not relevant. The only way in which it might be relevant is if interpreters were trained to gesture in certain ways. In fact, the interpreting tradition in the cultural context considered here (at the university in Russia where these interpreters were trained) often advises simultaneous interpreters to sit still at their desk, usually with hands folded, not gesturing, so as not to attract attention to themselves. Therefore, in the analysis of the results below, we will focus on the footing of the principal, giving special attention to the questions mentioned at the end of Section 1 above.

4.1 Gestures with referential functions

4.1.1 Representational gestures

Only 6% ($N=216$) of the 3,719 gestures in total were representational in function. This amount hardly differed in relation to the direction of interpreting between L1 and L2 (L1 to L2 7%, L2 to L1 6%). Examples of representational gestures included instances where the interpreter was rendering a number, and while doing so, traced the written number on the desk with their finger, as if their finger were a pencil writing the number. This helped one interpreter, for example, who was rendering a phrase from Russian into German: having heard in Russian "*ot soroka tysjač do semidesjati tysjač vidov v god*" ('from forty thousand to seventy thousand species per year') and saying in German, "*mm […] von… vierzig [bis ehm… seib]zig Tausend Arten per Jahr*" ('mm […] from… forty [to uhm… seven]ty thousand species per year'). (Square brackets will be used to indicate the speech or pauses co-occurring with the stroke phase of each gesture. The length of pauses relative to the speaker's speech rate is indicated by either two dots for a shorter pause or three dots or more for a longer pause.) With her two hands resting on the desk, the interpreter moves her right hand palm down and with the middle finger she "writes" the numbers 4 and 7 on the desk at the first and second moments enclosed in square brackets in the phrase transcribed above. This kind of representation of the content being interpreted appears to be a means for the interpreter to keep track of the information to be rendered. It is well known that numbers can present a challenge in simultaneous interpreting (Mazza, 2001; Pellatt, 2006) because the information they convey is not predictable in the way that, for example, fixed phrases in a language are. Indeed, it is common practice to write down numbers during simultaneous interpreting so as not to forget them. In this respect, it seems clear that the principal behind the gestures in this case is the interpreter herself; that is: it is far less plausible that the interpreter may have imagined the original lecturer himself writing down the numbers 4 and 7 when he was uttering "forty thousand" and "seventy thousand" during his lecture.

In another case, the interpreter renders a phrase about the famously extinct dodo bird, saying in Russian, "*i my vse znaem, [kak vygljadel dodo] v ètoj knižke*" ('and we all know [what the dodo looked like] in that book'). The gesture stroke co-occurring with the bracketed phrase involved holding up both open hands facing each other and quickly moving them downward in a wobbly path, as indicated in Figure 1.

This qualifies as a molding gesture, as if touching the surface of a medium-sized object that is somewhat taller than it is wide—just as

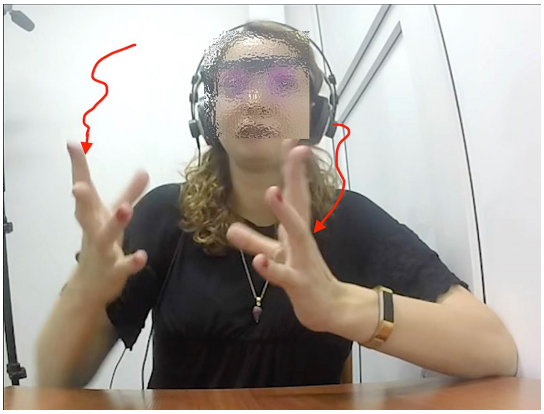


FIGURE 1
A molding representational gesture while saying in Russian, "*i my vse znaem, [kak vygljadel dodo] v ètoj knižke*" ('and we all know [what the dodo looked like] in that book').

the image of the extinct bird is often shown, in a standing position.[7] Here the principal of the gesture could be the interpreter, based on her own thinking for speaking, but it is also not implausible to imagine the original lecturer perhaps gesturing a rough image of a medium-sized dodo-object in the air as he mentioned it. Given that the interpreters were only hearing the lecturer and not seeing him, it is possible that they might mentally simulate (Marghetis and Bergen, 2014) the gestural production of the speaker they were hearing. Perhaps both phenomena are possible at the same time—the interpreter gesturing the general shape and size of her mental image of the dodo, enacting what one might plausibly imagine the lecturer could have done. The ambiguity here reflects one way of understanding the laminated nature of interpreters in their task as speakers (Goodwin and Goodwin, 2004).

4.1.2 Deictic gestures

An even smaller amount of the gestures, namely 3% ($N = 114$), was deictic in function (L1 to L2 4%, L2 to L1 2%). One type of deixis observed was that of pointing gestures, with an extended index finger or flat hand. However, the pointing was not to physical referents in the interpreters' surroundings; they were not talking about the place in which they were located while performing the task. Instead, in a few cases they pointed to a space off to the side when interpreting an utterance by the lecturer that made reference to a graph or map that he was showing. Again, the interpreter was not viewing a video of the speaker, and so had no information as to where the image being referred to was being shown (e.g., on which side of the speaker). In the example shown in Figure 2A, the speaker, interpreting the German lecture, says in Russian, "*[vot tut], …vy vidite, naprimer… u[tra]tu ploščadej doždevyx lesov*" ('[right here], … you see, for example… the [loss] of acreage of rain forests') and points to the upper right (Figure 2A), also directing his eye gaze there on "vot tut" ('right here') and also moves his right hand, fingers extended, in an arc to the right and makes a beat downward when saying "utratu"

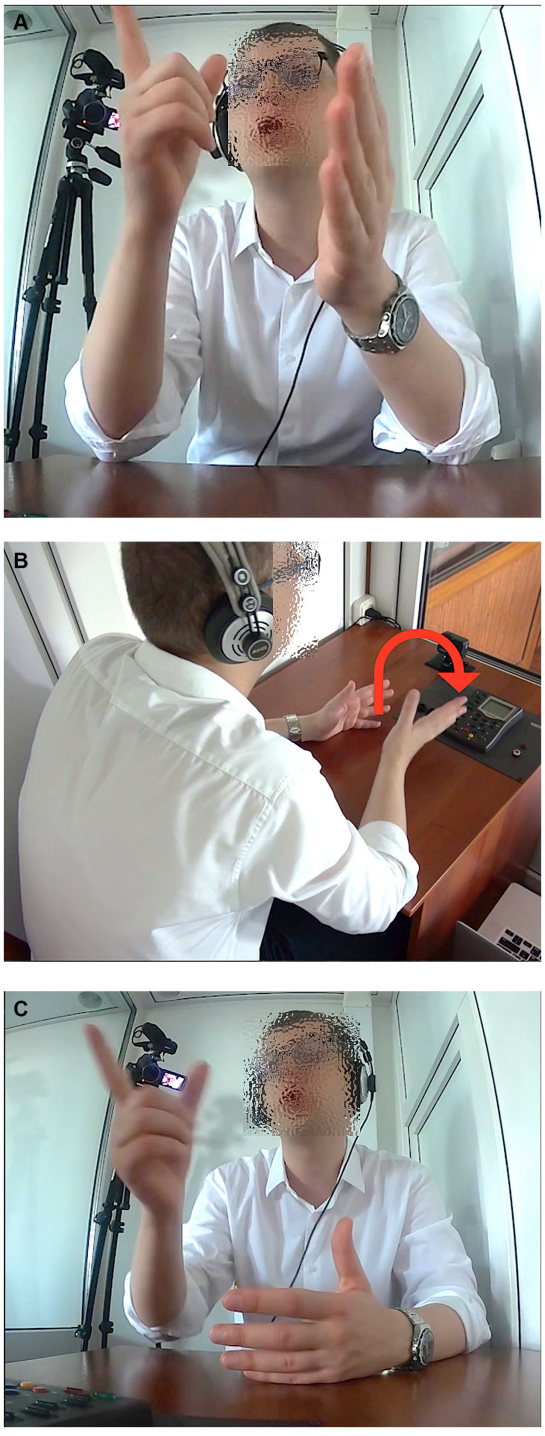


FIGURE 2
(A) A deictic gesture while saying in Russian, "*[vot tut], …vy vidite, naprimer…*" ('[right here], … you see, for example…). **(B)** A deictic gesture while saying in Russian, "*u[tra]tu ploščadej doždevyx lesov*" (the [loss] of acreage of rain forests'). **(C)** A deictic gesture while saying in Russian, "*no ne [tol']ko tam. V vostočnoj Azii ètі processy takže otmečajutsja*" ('but not [on]ly there. These processes have also been noted in East Asia').

7 See, for example, the well-known image of the dodo painted by George Edwards in 1626. https://en.wikipedia.org/wiki/Dodo#Contemporary_depictions.

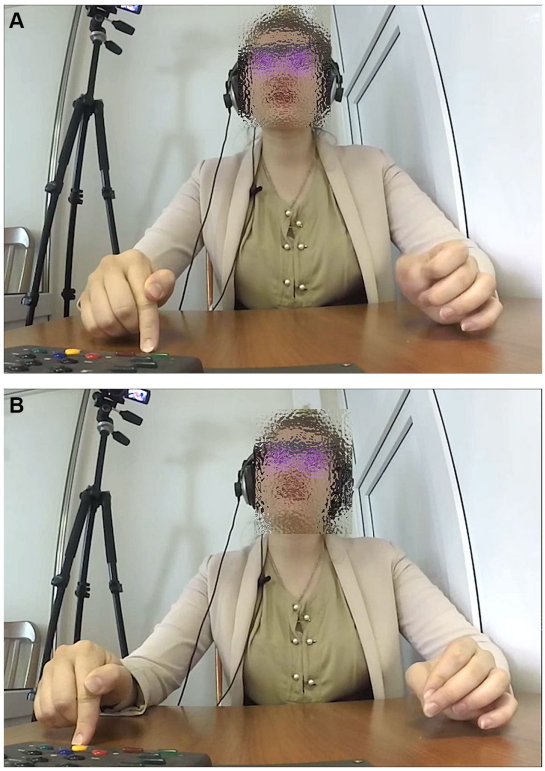


FIGURE 3
(A) A deictic gesture while saying in English, *"from th- [for]ty thousand."* **(B)** A deictic gesture while saying in English, *"to [se]venty thousand… species per year"*.

('loss') (Figure 2B), with his hand almost touching the desk.[8] Seconds later he says, "*no ne [tol']ko tam. V vostočnoj Azii èti processy takže otmečajutsja*" ('but not [on]ly there. These processes have also been noted in East Asia') and again makes a pointing gesture to the upper right with his index finger (Figure 2C).

The pointing was from the imagined viewpoint of the lecturer, pointing to the imagined physical chart being cited. Here one can argue that the principal of the gesture is the lecturer (or lecturer as mentally simulated). The viewpoint (conceptually, and even physically, as the interpreter looks up to the space he is pointing to) of the lecturer is blended with the interpreter's embodied rendering of it, as if the interpreter were pointing in place of the lecturer.

One other type of deixis involved touching. In the same portion of the Russian lecture mentioned earlier, when two numbers were cited, the interpreter, in this case rendering the lecture in English, said, "*from th- [for]ty thousand to [se]venty thousand… species per year*." During the bracketed syllables, she touched the desk in front of her in two different places, shown in Figures 3A,B, locating the amounts as points in space, metaphorically objectifying the quantities as locations. Interestingly, the second point, identifying a higher number, was laterally to the right of the first point deictically touched, as if on a number line.

8 The interpreter then makes additional pragmatic gestures for emphasis during this utterance, not indicated here.

As with the example of tracing the numbers on the desk, the gesture appears to be based on the interpreter as the principal, keeping track of the two quantities mentioned. It is also conceivable (though perhaps less plausible) that the interpreter may have gestured in this way based on an imagined (mentally simulated) anticipation of what the original speaker may have done when mentioning these numbers. This could constitute another example of the potential ambiguity of principal footing in the gesturing of the simultaneous interpreters.

Though the referential function of gesture (representation or deixis) constituted the smallest proportion of gestures used, it nevertheless raises some intriguing questions about the footing behind the interpreters' gestures. The use of gestures with pragmatic functions, discussed in the next section, presents further puzzles when the issue of stance comes to the fore.

4.1.3 Gestures with pragmatic functions

Forty four percent ($N = 1{,}638$) of the gestures were pragmatic in function. Here there was a small difference based on the direction of interpreting: L1 to L2 40%, L2 to L1 47%. Let us consider three examples of different kinds of pragmatic uses of gestures that were observed before considering how they might be interpreted in terms of footing. Though one specific instance of each will be described, each type was used by several of the interpreters.

In one example of interpreting from Russian to English, the participant uttered the phrase "<inbreath>… *[sev]en… [spe]cies [of] [birds]*" (followed by the phrase "*are now extinct there*") while holding her left hand in a position with the tips of the thumb and index finger touching, the fingers thus making a ring shape (even if not a perfect circle) while the other fingers were extended and slightly curved, as shown in Figure 4. She moved her hand down in a beat with the prosodic stress on each of the syllables marked in square brackets above.

This handshape, with the palm facing the central gesture space, as shown in Figure 4, is a variant of the ring gesture (Kendon, 2004, pp. 238–247; Müller, 2014b). It can be used by speakers (at least of the European languages studied to date in relation to gesture use) when making a precise point. Morris (2002, pp. 78–79) therefore refers to it as the "precision grip," as the thumb and forefinger would be used to grasp and hold a tiny object (on this see also Calbris, 2011, pp. 21–22; Lempert, 2011). Therefore the use of this gesture can be related to both epistemic stance (showing the precise certainty of the information being uttered) and relevance stance (showing that this information is important, putting it in focus).

In another example, the interpreter had referred to 90% of the species on Earth, after which he continued, interpreting from Russian to German, "*{wir können |nichts sag|en, ob sie [aus}sterben oder nicht], aber*" ('{we can|not say anything| about whether they are [dy}ing out or not], but'). In addition to the square brackets [] indicating the words co-occurring with a manual gesture, the curly brackets {} indicate the phrase with which the interpreter rapidly shook his head with small movements back and forth, to the left and to the right, several times. The vertical pipes | | indicate the syllables on which he raised his right shoulder slightly, once on each syllable. During the phrase in square brackets, he turns out his two open hands, fingers outstretched, as shown in Figure 5, making beat movements downward on the four syllables marked here: "*áusstérben óder nícht*."

The opening of the hands combined with the downward beats is similar to the palm-up gesture speakers may produce when presenting

FIGURE 4
A pragmatic precision-grip gesture while saying in English, "<inbreath>... *[sev]en. [spe]cies [of] [birds]*".

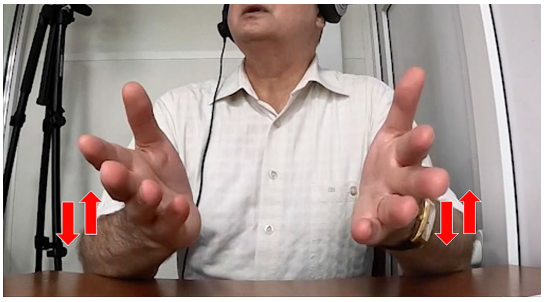


FIGURE 5
Pragmatic beat gestures while saying in German, *"áusstérben óder nícht"* ('are dying out or not').

a point (Müller, 2004). Combined with the lateral headshaking and the shoulder lifts, we see three components of the complex enactment known in English as a shrug. Debras (2017), Jehoul et al. (2017), and Streeck (2009, ch. 8) discuss the function of the shrug in expressing a stance less committed to the information being uttered (what Debras and Cienki, 2012, refer to as "dis-stance"), and/or uncertainty, in terms of epistemic stance. Even the various individual components may relate more to the expression of particular aspects of the stance. Thus while the lateral headshake is known to express negative assessment in most European cultures (Harrison, 2014), Debras (2017) notes that the raising of one shoulder more often expresses an affective stance (indifference or rejection), while the turning out (supination) of one or both hands correlates more with the attitudinal expression of incapacity to know or to take action. In the example considered here, the differential timing of the use of the different components shows a dynamic shift from negation (we cannot say whether these species will die out) to the attitudinal stance of admitting that we are incapable of knowing this.

The third example encompasses a set of instances of mentioning a point accompanied by a small turn out of the hand (resulting from a small rotation of the forearm), outward and back in, as in Figure 6A, or even just an extension of one or more fingers of the hand and then a return back to the starting position. In some instances it simply involved a lifting of one or two thumbs if the hands were folded on the desk, as shown in Figure 6B.

These types of gestures have been analyzed (Cienki, 2021) as miniature variants of the palm-up open hand, well known as a gesture used when presenting a point (Müller, 2004; Cooperrider et al., 2018). They provide a minimally effortful way to show information status (relevance stance), indicating the point being uttered verbally as something to be taken into consideration.

In terms of the principal footing, the question we are left with is whether the use of these stance-taking gestures in the context of simultaneous interpreting (where the original speaker is not seen) derives from the imagined behavior of the original lecturer (the imagined stance of that speaker), or the interpreter's own stance, or whether there might be other explanations. This issue will be considered further in the Discussion and Conclusions, below. It is interesting to remember, though, that the interpreters did not have any visible audience that they were speaking to. While it is true that they were being recorded, the larger camera was out of their view (behind them) and the other camera was a small, unobtrusive GoPro on the desk in front of them, as can be seen in Figure 2B. In addition, previous research (e.g., Mol et al., 2011) has shown that speakers do not produce more gestures simply when a camera is present or because of other people being in the room, but rather when they knew an addressee could see them, namely in settings where they could see their addressee's eye gaze. Nevertheless, in the present study, many of the interpreters produced many gestures.

4.1.4 Gestures with an adaptive function

Of the gestures serving an adaptive function, the majority (44%, $N = 1{,}636$) were self-adapters, with a slight difference in the amount depending in the direction of interpreting (L1 to L2 46%, L2 to L1 42%). Only 3% ($N = 115$) were other-adapters (3% being the proportion for both L1 to L2 and L2 to L1 interpreting). Most of the self-adapters were sustained in nature: the position that many of the interpreters assumed, starting with their hands folded on the desk, afforded movements like rubbing one hand with the fingers of the other (see

FIGURE 6
(A) A pragmatic small hand turn-out while saying in Russian, "*èto bylo [vygodno]*" ('it was [advantageous]'). **(B)** A pragmatic lifting of the thumbs while saying in English, "*what [kind of] animals*".

FIGURE 7
A sustained self-adapter, rubbing one's finger while speaking (time elapsed: 8 s).

Figure 7), or moving one's hand down to rub one's arm. Sometimes this even took more extreme forms, involving pulling on one's own skin.

In terms of principal footing, it does not seem likely, while engaged in the cognitively intense task of simultaneous interpreting, that the interpreter was picturing the original speaker making such small sustained movements while lecturing. It is much more plausible that such movements serve the interpreters' own purposes of cognitive focussing, perhaps even self-soothing (Freedman, 1972) to relieve some of the stress of the task. Here the principal behind these movements is more clearly the interpreter. In this regard, we might say that the lamination discussed above comes apart momentarily when orienting inward, using self-adapters; the interpreter's footing in such moments is less multifaceted than when engaged with outwardly-oriented gestures that might embody what the lecturers could have been doing as part of giving their talk.

5 Discussion and conclusion

We see that interpreters are laminated speakers in more ways than just in their use of speech (as Vranjes and Brône, 2021, point out). However, there are varying degrees of differentiation as to whose stance and principal footing they are expressing bodily.

On the verbal level, the principal of what the interpreter is uttering is clear: it is the speaker of the source text, in our case: the original lecturer being heard. Only rarely are interpreters the principal of the words they utter; this can occur momentarily when they correct what they said and add "Excuse me" or the like in the target language. Here the switch in footing is discrete (excusing themselves in that moment), sandwiched between the renderings for which the lecturer of the source text is clearly the principal; that is, they are not asking the hearer to excuse the original lecturer.

However, we have seen that it is often not possible to clearly determine the principal footing behind simultaneous interpreters' gestures. This is quite different from the situation that Vranjes and Brône (2021) describe for consecutive interpreters, where the speakers of the source text are present as interactants along with the interpreter and the audience of the interpretation. In that context, eye gaze direction, head nods, and manual pointing gestures are sometimes used to indicate that the principal of an interpreted utterance is not the interperpreter him/herself but the original speaker, who is visibly present. Vranjes and Brône point out that the verbal attribution of the principal can sometimes be confusing for listeners during interpreting, given interpreters' convention of maintaining the original speaker's use of the first-person pronoun (i.e., it can be confusing that the "I" used the interpreter means someone else). But gesture use in consecutive dialogue interpreting can disambiguate that the interpreter, as the animator and author of the interpreted utterance, is not the principal. The difficulty in determining the principal as displayed in simultaneous interpreters' gestures is partly due to the fact that all of the interactants (speaker of the source text, interpreter, and audience of the interpretation in the target language) are not sharing attention in the same space where they can all see each other. There is not one framework for deixis to operate in, for example.

Ascertaining the principal behind the pragmatic gestures observed in this study presents a particular puzzle, as noted earlier, and relates to the condition of the participants (speaker, interpreter, listener) not sharing one interactive space, visually accessible to all. Though some of the pragmatic uses of gesture discussed here (components of the shrug and hand turnouts as presentation gestures) were considered interactional functions by Bavelas et al. (1992), the interaction here is fictive, in the sense of Pascual (2002, 2014). Pascual builds on Talmy's notion of fictivity, which refers to "the imaginal capacity of cognition" (Talmy, 2000, p. 100) and an "as if" state of affairs. Therefore, fictive interaction can be distinguished from factual (objectively verifiable interaction with someone else in real time) and from fictional or fictitious [interaction "conceptualized as occurring in a fantasy world or even in a hypothetical or counterfactual scenario" (Pascual, 2006, p. 384)]. So while the fictive interaction in our study could theoretically be conceived of as mentally simulated interaction (e.g., with an imagined audience), the level of cognitive load that simultaneous interpreters are already handling makes this explanation less plausible.

A more tenable explanation for such use of pragmatic gestures might be that they are so ingrained as part of the process of spoken interaction, at least among adult speakers, that interpreters cannot help but produce them when they would themselves engage in stance-taking when presenting the points that they are uttering. This might explain the slightly greater proportion of pragmatic gestures when interpreting from L2 to L1. In one's native language and culture, one's routines for engaging in talk in interaction are more ingrained; one has a handy repertoire of recurrent gestures that one can resort to. (We can contrast this with the slightly higher proportion of self-adapters found in the interpreting from L1 to L2.) The fact of gesture use even in contexts in which no interlocutor is present or visible highlights the inherently intersubjective nature of language. As Cuffari (2024, p. 611) captures it, "Gesturing and intersubjectivity are multifaceted yet reciprocally informing phenomena that presuppose each other." However, Hostetter and Alibali (2008) point out that speakers do not gesture with every utterance, but rather they do so when the motivation to gesture reaches and exceeds a certain threshold. The threshold can be higher or lower depending on a complex of factors, including the individual speaker's habits, the cognitive effort they are exerting in the moment, the discourse context (what was being talked about previously), the social context (more formal versus more relaxed), etc. Given the varying strength that the various factors may have in the present context, it would explain the wide variation across the interpreters (individual variation) in their use not only of pragmatic gestures but of gestures in general.

In addition, the difficulty in terms of attribution of the principal behind gestures in this context is partly a factor of the nature of gesture in general as a semiotic system. Gesture is arguably more dependent on speech and contextual information in most contexts than speech is dependent on gesture (viz. Kibrik and Molchanova, 2013). Gesture is generally underspecified in form in relation to function. If we take deictic gestures, for example, it is well known that someone observing them is dependent on context for determining the target that the gesturer may have intended with their pointing (Kendon, 2004, ch. 11; Kita, 2003; Talmy, 2017). If we think of representational gestures, the depiction involved is always metonymic (synechdocal), iconically showing only a part of some referent (Mittelberg and Waugh, 2009, 2014; Müller, 2014a). That is part of how the modes of representation function: they provide schematic imagery. Some of the representation may be based on schematizations of everyday actions, such as PUT IN, TAKE OUT, SIT, RUN, etc., what Zlatev (2005, 2007) has called mimetic schemas (see also Cienki, 2013; Zlatev, 2014). Other instances of gesture use may draw upon even more general patterns in our everyday experience, what Johnson (1987) has discussed as image schemas, such as CONTAINMENT, BALANCE, or PATH (Cienki, 2005). Turning to pragmatic gestures, they are sometimes produced in less effortful forms; speakers might not produce the full compound enactment of a shrug, mentioned earlier, but just a small part of it, with less effort. The presentation gesture—the archetype of which might be the magician presenting the result of a trick and exclaiming "Ta daa!" with a full turning out of a palm-up open hand—is more often produced in everyday conversation in reduced forms, with the hand not fully turned palm up, and perhaps with just a finger extended outward (Cienki, 2021). In these ways, schematic instantiations of pragmatic gestures are all that speakers produce in many instances. As Mark Turner (personal communication, cited in Cienki, 2017) phrased it, if we consider any expression, be it verbal or gestural, "the product is a given precipitation of a process," with the process being the conceptualization in the given context that led to how the expression was formulated. This can be more or less elaborate (more or less schematic and metonymic) for any verbal or gestural expression.

In terms of limitations of the study, we acknowledge that the setting was not completely authentic. It was not a live lecture being interpreted for an audience that was visibly present. This was a factor of wanting to have the interpreters only hear the lectures without being influenced by seeing the original speakers' gestures; there was also the logistical factor of wanting several dozen interpreters to interpret the same lecture and the logistical and scheduling challenges that would have arisen if we had had to bring in an audience for each interpreting session. An extension of this project will have interpreters view the video-recordings of the lectures that they are interpreting, bringing the study closer to authentic conditions, particularly those used for interpreting in videoconferences. The study also faced difficulties arising from the COVID-19 pandemic, namely that the lockdown restrictions meant that interpreters could only participate during certain time periods. This resulted in a somewhat larger number of participants for the Russian-English study than for the Russian-German study, as the data collection for the latter overlapped with the pandemic.

In conclusion, we see that the difference in status between the semiotic systems of lexico-grammar versus manual gesture plays out in terms of the difference in how principal footing can be attributed on the verbal level in simultaneous interpreting and in terms of the use of gesture. Looking at gesture, we see in many cases the lamination and the ambiguity of the principal footing. This schematicity and ambiguity of gesture may be part and parcel of what is involved in interpreters' thinking for speaking. In particular, their frequent use of pragmatic gestures plays on the border between what the original lecturer may have done when expressing a stance toward the topic mentioned verbally and the interpreter's own stance.

McNeill (2000, 2013) argues that gesture provides a window onto the mind. Through the window of simultaneous interpreters' gestures, we can catch glimpses into the blending of viewpoints that

thinking for simultaneous interpreting appears to involve (as per Cienki and Iriskhanova, 2020) in different ways, changing over time. Such thinking for interpreting clearly differs from the process of unpacking one's own idea units that McNeill argues takes place in spontaneously expressing one's own thoughts. The ideas to be spoken are presented to interpreters in the utterances in the source language, rather than arising from their own personal engagement in thinking and interacting with others. As a re-presenter of the original lecturer's words in another language, interpreters might project what such a lecturer might have done in the context of presenting the given ideas, but they also surely incorporate elements of their own repertoire of how they speak and present ideas, including gesturally.

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 humans were approved by the Research Ethics Committee of the Science and Technology Board at Moscow State Linguistic University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

AC: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Data collection and annotation were supported by Russian Science Foundation grant no. 19-18-00357 and the development of procedures for analyzing stance-taking gestures was supported by RSF grant no. 24-18-00587—both grants awarded for research carried out at Moscow State Linguistic University.

Acknowledgments

Credit for the data collection, annotation, and analysis is due to the following members of the PoliMod Lab in the Centre for Socio-Cognitive Discourse Studies (SCoDis); in alphabetical order: Olga Agafonova, Olga Iriskhanova, Varvara Kharitonova, Anna Leonteva, Alina Makoveeva, Andrej Petrov, Olga Prokofjeva, Evgenija Smirnova, Maria Tomskaja, and also Geert Brône. The article arose out of a talk co-authored with Geert Brône for the 2022 conference of the International Society for Gesture studies (ISGS) in Chicago, IL.

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

AIIC (2004/2012). Practical guide for professional interpreters. Geneva: AIIC.

Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse Process.* 15, 469–489. doi: 10.1080/01638539209544823

Bavelas, J., Gerwing, J., Sutton, C., and Prevost, D. (2008). Gesturing on the telephone: independent effects of dialogue and visibility. *J. Mem. Lang.* 58, 495–520. doi: 10.1016/j.jml.2007.02.004

Biber, D., and Finegan, E. (1988). Adverbial stance types in English. *Discourse Process.* 11, 1–34. doi: 10.1080/01638538809544689

Biber, D., and Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9, 93–124.

Bressem, J., Ladewig, S. H., and Müller, C. (2013). "Linguistic annotation system for gestures (LASG)" in Body—Language—Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Teßendorf, vol. *1* (Berlin: De Gruyter Mouton), 1098–1124.

Bressem, J., and Müller, C. (2014). "A repertoire of recurrent gestures of German" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1575–1591.

Calbris, G. (1990). The semiotics of French gestures. Bloomington, IN: Indiana University Press.

Calbris, G. (2011). Elements of meaning in gesture. Amsterdam: John Benjamins.

Cienki, A. (2005) in "Image schemas and gesture" in from perception to meaning: Image schemas in cognitive linguistics. ed. B. Hampe (Berlin: Mouton de Gruyter), 421–441.

Cienki, A. (2010). "Multimodal metaphor analysis" in Metaphor analysis: research practice in applied linguistics, social sciences and the humanities. eds. L. Cameron and R. Maslen (London: Equinox), 195–214.

Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Rev. Cogn. Linguist.* 11, 417–432. doi: 10.1075/rcl.11.2.13cie

Cienki, A. (2017). Utterance construction grammar (UCxG) and the variable multimodality of constructions. *Linguist. Vanguard* 3:48. doi: 10.1515/lingvan-2016-0048

Cienki, A. (2021). From the finger lift to the palm-up open hand when presenting a point: a methodological exploration of forms and functions. *Lang. Modalities* 1, 17–30. doi: 10.3897/lamo.1.68914

Cienki, A., and Iriskhanova, O. K. (2020). Patterns of multimodal behavior under cognitive load: an analysis of simultaneous interpretation from L2 to L1. *Voprosy Kognitivnoy Lingvistiki* 1, 5–11. doi: 10.20916/1812-3228-2020-1-5-11

Cooperrider, K., Abner, N., and Goldin-Meadow, S. (2018). The palm-up puzzle: meanings and origins of a widespread form in gesture and sign. *Front. Commun.* 3:23. doi: 10.3389/fcomm.2018.00023

Cuffari, E. C. (2024). "Gesture and intersubjectivity" in The Cambridge handbook of gesture studies. ed. A. Cienki (Cambridge: Cambridge University Press), 599–615.

Darò, V., and Fabbro, F. (1994). Verbal memory during simultaneous interpretation. Effects on the phonological loop. *Appl. Linguis.* 15, 365–381. doi: 10.1093/APPLIN/15.4.365

Debras, C. (2017). The shrug: forms and meanings of a compound enactment. *Gesture* 16, 1–34. doi: 10.1075/gest.16.1.01deb

Debras, C., and Cienki, A. (2012). Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking. International proceedings of the ASE 2012 international conference of social computing, Amsterdam, Netherlands, 3–5 September 2012.

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi: 10.1515/9783110880021.57

Freedman, N. (1972). "The analysis of movement behavior during the clinical interview" in Studies in dyadic communication. eds. A. W. Siegman and B. Pope (New York, NY: Pergamon Press), 153–175.

Gernsbacher, M. A., and Schlesinger, M. (1997). The proposed role of suppression in simultaneous interpretation. *Interpreting* 2, 119–140. doi: 10.1075/intp.2.1-2.05ger

Goffman, E. (1981). Forms of talk. Philadelphia, PA: University of Pennsylvania Press.

Goodwin, C., and Goodwin, M. H. (2004) in "Participation" in a companion to linguistic anthropology. ed. A. Duranti (Oxford: Basil Blackwell), 222–243.

Grishina, E. A. (2017). Russkaja žestikuljatsija s lingvisticheskoj tochki zrenija: Korpusnye issledovanija [Russian gesticulation from a linguistic point of view: Corpus studies]. Moscow: Languages of Slavic Culture.

Harrison, S. (2014). "Head shakes: variation in form, function, and cultural distribution of a head movement related to 'no'" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1496–1501.

Harrison, S. (2021). The feel of a recurrent gesture: embedding the vertical palm within a gift-giving episode in China (aka the 'seesaw battle'). *Gesture* 20, 254–284. doi: 10.1075/gest.21003.har

Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495

Jehoul, A., Brône, G., and Feyaerts, K. (2017). The shrug as marker of obviousness: corpus evidence from Dutch face-to-face conversations. *Linguist. Vanguard* 3, 1–9. doi: 10.1515/lingvan-2016-0082

Johnson, M. (1987). The body in the mind: the bodily basis of meaning, imagination, and reason. Chicago, IL: University of Chicago Press.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *J. Pragmat.* 23, 247–279. doi: 10.1016/0378-2166(94)00037-F

Kendon, A. (2004). Gesture: visible action as utterance. Cambridge: Cambridge University Press.

Kendon, A. (2017). Pragmatic functions of gestures: some observations on the history of their study and their nature. *Gesture* 16, 157–175. doi: 10.1075/gest.16.2.01ken

Kibrik, A. A., and Molchanova, N. B. (2013). Channels of multimodal communication: relative contributions to discourse understanding. Proceedings of the 35th annual conference of the cognitive science society, Berlin, Germany, 2704–2709.

Kita, S. (2003). Pointing: where language, culture, and cognition meet. Mahwah, NJ: Lawrence Erlbaum.

Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059

Kok, K., Bergmann, K., Cienki, A., and Kopp, S. (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture* 15, 37–59. doi: 10.1075/gest.15.1.02kok

Ladewig, S. H. (2014). "Recurrent gestures" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1558–1574.

Langacker, R. (1988). "A usage-based model" in Topics in cognitive linguistics. ed. B. Rudzka-Ostyn (Amsterdam: John Benjamins), 127–161.

Lempert, M. (2011). Barack Obama, being sharp: indexical order in the pragmatics of precision-grip gesture. *Gesture* 11, 241–270. doi: 10.1075/gest.11.3.01lem

Marghetis, T., and Bergen, B. K. (2014). "Embodied meaning, inside and out: the coupling of gesture and mental simulation" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 2000–2007.

Martín de León, C., and Fernández Santana, A. (2021). Embodied cognition in the booth: referential and pragmatic gestures in simultaneous interpreting. *Cogn. Linguist. Stud.* 8, 277–306. doi: 10.1075/cogls.00079.mar

Mazza, C. (2001). Numbers in simultaneous interpreting. *Interpreters' Newsletter* 11, 87–104.

McNeill, D. (1992). Hand and mind: what gestures reveal about thought. Chicago, IL: Chicago University Press.

McNeill, D. (2000). Language and gesture. Cambridge: Cambridge University Press.

McNeill, D., and Duncan, S. (2000). "Growth points in thinking-for-speaking" in *Language and gesture.* Ed. D. McNeill (Cambridge: Cambridge University Press), 141–161.

McNeill, D. (2013). "Gesture as a window onto mind and brain, and the relationship to linguistic relativity and ontogenesis" in Body – language – communication. eds. C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill and S. Teßendorf, vol. *1* (Berlin: De Gruyter Mouton), 28–54.

Mittelberg, I., and Waugh, L. R. (2009). "Metonymy first, metaphor second: a cognitive-semiotic approach to multimodal figures of thought in co-speech gesture" in Multimodal metaphor. eds. C. Forceville and E. Urios-Aparisi (Berlin: Mouton de Gruyter), 329–356.

Mittelberg, I., and Waugh, L. (2014). "Gestures and metonymy" in Body – language – communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1747–1766.

Mol, L., Krahmer, E., Maes, A., and Swerts, M. (2011). Seeing and being seen: the effects on gesture production. *J. Comput. Mediat. Commun.* 17, 77–100. doi: 10.1111/j.1083-6101.2011.01558.x

Morris, D. (2002). Peoplewatching. London: Vintage (Original work published 1977 as Manwatching. London: Jonathan Cape).

Müller, C. (1998). Redebegleitende Gesten. Kulturgeschichte–Theorie–Sprachvergleich. Berlin: Berlin Verlag A. Spitz.

Müller, C. (2004). "Forms and uses of the palm up open hand: a case of a gesture family?" in The semantics and pragmatics of everyday gestures. eds. C. Müller and R. Posner (Berlin: Weidler), 233–256.

Müller, C. (2014a). "Gestural modes of representation as techniques of depiction" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1687–1701.

Müller, C. (2014b). "Ring-gestures across cultures and times: dimensions and variation" in Body – language – communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1511–1522.

Pascual, E. (2002). Imaginary trialogues: conceptual blending and fictive interaction in criminal courts. Utrecht: LOT Dissertation Series.

Pascual, E. (2006). Questions in legal monologues: fictive interaction as argumentative strategy in a murder trial. *Text Talk* 26, 383–402. doi: 10.1515/TEXT.2006.014a

Pascual, E. (2014). Fictive interaction: the conversation frame in thought, language, and discourse. Amsterdam: John Benjamins.

Pellatt, V. (2006). "The trouble with numbers: how linguistic, arithmetical and contextual complexity affect the interpretation of numbers" in Professionalization in interpreting: International experience and development in China. eds. M. Chai and J. Zhang (Shanghai: Foreign Language Education Press), 350–365.

Poggi, I. (2014). "Semantics and pragmatics of symbolic gestures" in Body – Language – Communication. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *2* (Berlin: De Gruyter Mouton), 1481–1496.

Setton, R., and Dawrant, A. (2016). Conference interpreting: A trainer's guide. Amsterdam: John Benjamins.

Slobin, D. I. (1987). "Thinking for speaking" in Proceedings of the 13th annual meeting of the Berkeley linguistics society. eds. J. Aske, N. Beery, L. Michaelis and H. Filip (Berkeley, CA: Berkeley Linguistics Society), 435–445.

Slobin, D. I. (1996). "From 'thought and language' to 'thinking for speaking'" in Rethinking linguistic relativity. eds. J. Gumperz and S. C. Levinson (Cambridge: Cambridge University Press), 70–96.

Sloetjes, H., and Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. Available at: https://aclanthology.org/volumes/L08-1/

Stachowiak-Szymczak, K. (2019). Eye movements and gestures in simultaneous and consecutive interpreting. Cham: Springer.

Stelma, J. H., and Cameron, L. J. (2007). Intonation units in spoken interaction: developing transcription skills. *Text Talk* 27, 361–393. doi: 10.1515/TEXT.2007.015

Streeck, J. (2009). Gesturecraft: The manu-facture of meaning. Amsterdam: John Benjamins.

Talmy, L. (2000). Toward a cognitive semantics. Volume 1: Concept structuring systems. Cambridge, MA: The MIT Press.

Talmy, L. (2017). The targeting system of language. Cambridge, MA: The MIT Press.

Vranjes, J., and Brône, G. (2021). Interpreters as laminated speakers: gaze and gesture as interpersonal deixis in consecutive dialogue interpreting. *J. Pragmat.* 181, 83–99. doi: 10.1016/j.pragma.2021.05.008

Will, I. (2022). Recurrent gestures of Hausa speakers. Leiden: Brill.

Zlatev, J. (2005). "What's in a schema? Bodily mimesis and the grounding of language" in From perception to meaning: Image schemas in cognitive linguistics. ed. B. Hampe (Berlin: Mouton de Gruyter), 313–342.

Zlatev, J. (2007). "Language, embodiment and mimesis" in Body, language, mind. Vol 1: Embodiment. eds. T. Ziemke, J. Zlatev and R. Frank (Berlin: de Gruyter), 297–337.

Zlatev, J. (2014). Image schemas, mimetic schemas and children's gestures. *Cogn. Semiotics* 7, 3–29. doi: 10.1515/cogsem-2014-0002

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 October 2024
DOI 10.3389/fpsyg.2024.1436191

Check for updates

OPEN ACCESS

EDITED BY
Geert Brône,
KU Leuven, Belgium

REVIEWED BY
Costanza Navarretta,
University of Copenhagen, Denmark
Isabella Poggi,
Roma Tre University, Italy

*CORRESPONDENCE
Rebecca Clift
✉ rclift@essex.ac.uk

RECEIVED 21 May 2024
ACCEPTED 02 September 2024
PUBLISHED 28 October 2024

CITATION
Clift R (2024) What is in a look? The accountability of gaze in trajectories to conflict.
Front. Psychol. 15:1436191.
doi: 10.3389/fpsyg.2024.1436191

COPYRIGHT
© 2024 Clift. 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.

What is in a look? The accountability of gaze in trajectories to conflict

Rebecca Clift*

Department of Language and Linguistics, University of Essex, Colchester, United Kingdom

This study investigates the role of gaze in initiating episodes of conflict by examining, using multimodal conversation analysis, a set of cases in which a recipient is prompted to speak by another's extended gaze. In these cases, this recipient response may be, e.g., "What," or a more elaborate demand for an account, such as "Why are you looking at me like that for?" Here we investigate the characteristics of the gaze that prompts such responses, and what actions such responses constitute. While "What" compositionally resembles other-initiated repair, its sequential position characterizes it as a so-called "go-ahead" action. In these cases, the sequential positioning of such gazes, constituting it structurally as a so-called "pre," alongside its durational characteristics and facial expression, are examined to identify the normative associations of gaze and subsequent conduct that make such gazes accountable.

KEYWORDS

gaze, conflict, staring, accountability, other-initiated repair, presequences, conversation analysis

1 Introduction

From Darwin (1872) onwards, there has been a recognition that a particular form of fixed or studied eye-gaze—a so-called "stare"—in specific contexts may be associated with hostility.[1] At its extreme, such hostility is perhaps most acutely embodied in the phenomenon of the so-called "hate stare" leveled at African-Americans by White Americans. The white journalist, John Howard Griffin, passing as an African American man in the segregationist south of the US in the 1950s, captures one such occurrence:

1 This article was written while I was in receipt of a Distinguished Scholar Award from the Harry Frank Guggenheim Foundation for the project "Behind Closed Doors: Trajectories to Violence in Intimate Interaction," and I gratefully acknowledge the support of the Foundation, and of the Arts and Humanities Impact Acceleration Fund at the University of Essex in undertaking this research. I have shown the data in various presentations, among them, at Teacher's College, Columbia University, New York, Language and Social Interaction working group, in 2018; at Huddersfield University in 2019; at Uppsala University, for the Transition and Sustainability of Communicative Competencies in Interactions Involving Young People with Communication Disabilities (TRANSCOMM) Workshop 2: Tensions, Troubles and Displays of Affect, in 2020; Ocean University, Qingdao, Shandong, China, and the Dutch Linguistic Society, Dutch Society for Applied Linguistics and LOT, Netherlands Graduate School, in 2021; and the University of Salerno in 2022. I am grateful to the audiences at these institutions, and the reviewers and editors of this article, for their constructive comments and feedback.

FIGURE 1
Emily's studied gaze at Tom, extract 2, l.5. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

> It came from a middle-aged, heavy-set, well-dressed white man. He sat a few yards away, fixing his eyes on me. Nothing can describe the withering horror of this. You feel lost, sick at heart before such unmasked hatred, not so much because it threatens you as because it shows humans in such an inhuman light. You see a kind of insanity, something so obscene the very obscenity of it (rather than its threat) terrifies you (Griffin, 1961).

Griffin's account leaves no doubt as to the status of such gazes as actions. In addition, on a more mundane level, perceiving oneself to be stared at in a public space is one that many will have experienced. Robert De Niro's famous line from "Taxi Driver"—"You lookin' at me?"—captures the response to the perceived threat conveyed by an extended eye gaze. Such instances testify to the culturally salient associations between extended gaze and potential hostility, an association picked up by Goffman, who, on observing brief moments between strangers in public spaces, anchors this hostility in the "invasion of informational preserve" (Goffman, 1971, p. 54) that staring represents.

What follows contrasts with Goffman's observations of interactions between strangers in public by examining cases in domestic settings in which a party responds verbally to an extended gaze by another, who is an intimate—in our cases, a family member. In so doing, we aim to establish the interactional implications of the studied gaze and the ways in which the hostility apparent when it is deployed between strangers is played out in its use between intimates.

The data here are taken from corpora of filmed family interaction. Most were taken from a corpus, edited parts of which were originally broadcast on Channel 4 (UK) from 2008 to 2009 as part of the TV series "The Family." The broadcast programs were taken from approximately 1,500 tapes of two British families filmed continuously in their homes across 100 days by over 20 cameras. One family can be seen in extracts 2, 4, 8, 9, 10, and 11; the other can be seen in extract 1. Although the whole corpus comprised a whole week's worth of raw unedited footage of one family, as well as the broadcast footage of both families, all the extracts transcribed here are taken from the broadcast footage.[2] Extract 7 was taken from a corpus of informal family interactions recorded by the participants themselves.[3]

Figure 1 shows one instance of the phenomenon of interest. It captures the studied gaze of over 3 s by a teenage girl, Emily, across the dining table to her younger brother, Tom—an action that initiates a trajectory leading Emily to abandon her meal and go to her bedroom in the midst of the ongoing interactional conflict.

The focus in such cases is how such gazes are treated interactionally by both parties—that is, their sequential implications. As we shall see, it is overwhelmingly the case that such gazes, such as the one above, are taken to adumbrate some kind of negative or even hostile stance and initiate a trajectory of interactional conflict.

Of course, experience tells us that in other contexts (such as between lovers or intimates generally, for example), the studied mutual gaze may also be deployed to more affiliative ends. We return to such alternative possibilities in due course, but in the corpus of hundreds of hours' worth of filmed family interaction examined for this research, it was the case that *none* of the instances of studied eye-gaze that we encountered were in such highly affiliative environments.[4] This empirical skewing, reflected in the data examined here, suggests a strong preference for the studied gaze to be taken as potentially hostile. We examine this phenomenon in what follows. We initially sketch relevant work on gaze and embodiment, before discussing an initial instance of a held gaze as clearly problem-implicative. In examining a number of cases in analytic detail to track the interactional trajectory from the gaze initiation onwards, we then investigate what action the response to the gaze constitutes. This, in turn, illuminates the action that the held gaze itself is implementing. We show how the fixed gaze is taken to adumbrate a problem, before going on to examine some cases where this problem-implicativeness and potential source of conflict may be defused. In conclusion, we also discuss a clear exception to the hostility implied by a fixed gaze.

2 Background: work on gaze and embodiment

Over 60 years of psychological research has sought to elaborate on Darwin's original observations on gaze, from Gibson and Pick (1963), Argyle and Dean (1965), and Kendon (1967) onward; for a broad overview, see Hessels (2020). The affective possibilities of eye-gaze have been a consistent focus in such research (see, e.g., Green et al., 2003; Eisenbarth and Alpers, 2011; and for a review, see Hietanen, 2018). Work in multimodal Conversation Analysis (CA) has, since Goodwin (1979, 1981), sought to build on this study by bringing a consideration of sequential positioning (on which, see, e.g., Clift et al.,

2 The edited nature of the footage clearly raises issues for the analyst, but I hope it will be apparent that, as far as can be established, none of the extracts presented here have been analytically compromised.

3 I thank Jenny Mandelbaum for making these data available to me. I am also grateful to Dragonfly Productions for making "The Family" data—both edited for broadcast, and raw, unedited data—available.

4 This of course may be an artifact of the kinds of contexts video-recorded for conversation-analytic research (on which, see Clift, 2024).

2013) to the study of gaze and embodiment in interaction. Unlike much previous work on gaze which focuses on establishing the "meaning" of particular forms of gaze, CA examines actions and the practices that deliver actions across sequences, and uniquely takes the participants' systematic responses to those practices to be criterial for understanding a practice as displaying, for example, hostility or tenderness.

In a pioneering CA study of children from 1 to 2½ years of age, Kidwell (2005) shows how while engaging in activities such as biting, pushing, or hitting, children can look to their caregivers and differentiate between a "mere look" from them and a more extended, sanctioning gaze. What Kidwell calls "the look" is of relatively long duration, alights on a target or targets, and is produced as "an activity in its own right" (Kidwell, 2005, p. 429). These children are thus analyzing their own conduct: doing something sanctionable, then looking to see the ways in which they are being monitored. A central feature of "the look" is that the gaze is held. In this respect it resembles a number of other embodied phenomena characterized by so-called "holds": body posture (Sikveland and Ogden, 2012; Groeber and Pochon-Berger, 2013; Li, 2014; Floyd et al., 2016; Manrique, 2016), hand gestures (Clift, 2020), and facial expressions (Clift and Rossi, 2023) can all be used as well as gaze (Rossano, 2012) to indicate the at-that-moment unresolved status of a sequence and disruption to its progressivity (Schegloff, 2007, pp. 14–15)—and, as such, problem-implicative. All of this work has served to emphasize the importance of examining, not just the gaze itself, but its associated embodiment and the sequential environment in which these are produced for understanding its implications for action.

3 An initial instance: the held gaze as problem-implicative

In mundane interaction, the problem-implicativeness of a held gaze is nowhere more apparent than in the following instance. Here a young woman, Shay, voices her distress (l. 1-2), having just phoned her estranged mother, then registers her fiancé, Sunny, holding his gaze (see Figure 2).[5] She responds, after 3.5 s of a visible gaze from Sunny, with an apology to him (l.5) for what she takes to be the offense she has caused:

(1) Angry[6] (Clift F:2:3: 41–54)

Sha=Shay (eye gaze *); Sun=Sunny, Shay's fiancé (eye gaze +; embodiments •); Pol=Polly, Sunny's mother.

A display of distress might expect some kind of empathic response (see Heritage, 2011, on empathic moments)—particularly in the case of intimates—and Shay looks to Sunny in the wake of her turn (Figure 3). But as we see in Figure 2, Sunny meets her expression of distress with an impassive face and his head in a hold. As Figure 3 shows, he is sitting at a computer, and it is likely that he is focussing on the screen; in any case, he does not respond to Shay. Her apology

```
01  Sha:  ~All I'm thinking in the back of my head is,
02        if my Dad was here,~
03        +(3.5)+
04  sun:→+#gaze held, probably to computer, face impassive---> l.21
    Fig:   #2
05  Sha:  *'M sorry.
06        *-to Sun---> l.8
07        (1.2)
08  Sha:  Are you angry with me #at something I've said.=
    Fig:                        #3
09  Pol:  =No:[::!*
10  sha:  --------*
11  Sun:        [+'m not angry with anyone
12        (1.2)
13  Sun:  [I gotta shave.
14  Sha:  [Jis' look at yer face!
15        (1.0)
16  Sun:  That's my face. That's what you're marrying.
17        Get used [to it.
18  Pol:           [He he
19  Sha:  Heh heh.
20        (0.4)+
21  sun:  ---->+
22  Sun:  •Heh [heh. S(h)tuck with it for the rest of yer l(h)ife.•
23        •smiling broadly, meets Sha's gaze on 'S(h)tuck'-------- •
24  Pol:       [God.
```

and, in the face of further silence (l.7), her inquiry as to whether she has caused offense, suggest her understanding of his fixed gaze, impassive expression, and non-response collectively to project hostility: "are you angry with me at something I've said" (l.8). It is Poli, Sunny's mother, who immediately and most straightforwardly produces a prosodically emphatic denial (l.9) followed by Sunny, whose "I'm not angry with anyone" (l.11) constitutes a somewhat tepid rebuttal. Shay's response to this, "Jis' look at yer face," (l.14) provides a retrospective account of her apology and its rationale. It is this response—and its reference to "yer face"—that suggests Shay understands the held gaze, with the impassive face, to be projecting negativity—and specifically anger. Sunny's delivery of his subsequent response (ll.16–17), delivered in a brusque, mock-angry tone, plays on this misapprehension.

This exemplar thus makes plain the problem-implicativeness, if not the potential hostility, attached to a held gaze with an impassive face, although of course this instance was not directed to a target and was produced in a sequential position where some response might have been expectably due. In each of the cases that follow, we further explore the sequential implicativeness of the sustained gaze by examining instances which, in contrast, are directed at a specific target and which initiate a verbal sequence.

4 The gaze as action

The following instance, captured in Figure 1, takes place in the course of a family dinner. In the wake of an exchange between Jane,

5 Gazes are timed from the beginning of the gaze to the recipient's response. I am grateful to Dr. Nicole Smith for helping me with the timings of the gazes here in ELAN.

6 The transcriptions conform to the conventions developed by Lorenza Mondada and are set out at https://www.lorenzamondada.net/_files/ugd/ba0dbb_3978d2a34cf44376adb7a341975d23aa.pdf.

the mother, and her 14-year-old son, Tom, about his long fringe (l.1–3), 19-year-old Emily looks up from her meal and, with an impassive face, produces a sustained gaze of 4 s at Tom. In the course of the gaze, she suspends her cutlery in a hold so that it is evident that she is wholly preoccupied with looking; the "activity in its own right" observed by Kidwell (2005, p. 429). In response to the look, Tom, who was in the course of bringing a glass to his lips, halts the raise, tilts his head up to fix his gaze on Emily, and says: "What," l.7:

(2) Nothing

Jan=Jane, mother (gaze €, embodiments ‡); Emi=Emily, 19 year-old daughter (gaze *, embodiments •); Tom=Tom, 14 year-old son (gaze +, embodiments %)

```
01  Jan:   Doesn't that get on ‡yer nerves¿‡
                              ‡waves finger at own fringe‡
02         (1.8)
03  Tom:   It doe- no, not really, I can see straight throu(h)gh it.
04         (1.4)
05         •(3.0)•
    emi:   •lifts head to gaze at T; #holds gaze, spoon and fork
           suspended, while swallowing and licking lips•
    Fig:                              #1
06         +%(1.0)%+
    tom:   +  to E +
    tom:    %lifting glass to drink, halts%
07  Tom: → What.
08         (2.2)
09  Emi:   Nothing. I didn't say any‡€thing.
    jan:                            ‡stops chewing, holds-->l.13
    jan:                             € to E ---> l.15
11  Tom:   °Don't stare at me then°.
    jan:                        ---->‡
12         (.)
13  Jan:   ‡Stop being so agg[ressive Emily [(that face-)‡€
           ‡Raises finger and jabs at E on 'Stop'‡
                                                      -->€
14  Emi:                      [What?          [Actually Tom didn't
15         need to* have a-* *say anything to me earlier.*
                  *  to T  * *           to J              *
16         •                 (2.2)
    emi:   •gaze down to plate as she forks up pasta -->
17  Emi:   He was the one who was aggr•*e[ **ssive.*
                           --->         •* T **  J   *
18  Jan:                                   [STOP being so
           aggress[ive.
19  Emi:           [↑He was the one who was aggress[ive!
20  Tom:                                            [Course,
21         (-), yeah, course.
```

Emily's response (l.9), with its brusque denial that she said anything, itself receives a rebuke from Tom explicitly referencing the stare (l.11) and thus accounting for his verbal initiation of the sequence in l.7. It is at this point that Jane intervenes; her assessment of Emily's behavior as "so aggressive" (l.13) clearly treats Emily as the antagonist in this exchange—and one that leads, some turns later, to Emily abandoning her meal and going up to her bedroom. However, lls. 14-15 show that Emily's response treats an earlier exchange as the origin of the conflict, with Tom as the initiator. The hostility attributed to the direct gaze in this exchange, and subsequently embodied verbally by Emily, in contrast to that in extract (1), thus turns out to be warrantable—and initiates an episode of conflict in the here-and-now.

FIGURE 2
Sunny's gaze, extract 1, l.4. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

FIGURE 3
Shay's gaze to Sunny, extract 1, l.8. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

In attempting to identify the kind of action a studied gaze is taken to be implementing, the treatment by Emily of Tom's "What" is here critical. In insisting that she "did not say anything" (l.9)—something which is, of course, true—she treats it as an other-initiated repair. Her own subsequent "What?" at l.14 might appear, at first glance, to warrant this treatment of Tom's turn, as it initiates repair on his softly produced l.11.

Instances such as the one below might also appear to endorse an understanding of "What" as initiating repair. Here, Michael at l.4 initiates repair on Shane's "some" in l.1, and is completed by "saline solution" at l.6:

(3) Saline solution (Schegloff, 1997, p. 515)

Chicken Dinner, 48:34–49:11. Sha = Shane; Mic = Michael.

```
01  Sha:    =Lemme have some (0.2) t'night (.)
02           Lemme hev- cz I ran ou:t.
03           (0.4)
04  Mic:→   What.
05           (.)
06  Sha:    u-Saline solution gunnah git s'm
07           duhmorr'.
08           (0.7)
09  Mic:    Mm:.
```

However, as Schegloff et al. (1977) note, the repair is addressed to problems in the understanding of, and initiated on, *talk*—and so Tom's "What" as a response to Emily's look cannot be initiating repair. But another instance in which "What" responds to a gaze, and moreover, another such trajectory toward conflict, sheds light on this issue. In (4) below, a recipient—as it happens, Emily—responds to a studied gaze with "What." Earlier in the evening, Emily had been upbraided by her parents for consistently going out late at night to clubs and then calling in sick to work. The extract below takes place later that evening. Just beforehand, Simon, Emily's father, having opened the front door to a taxi driver whom Emily has evidently called, then enters Emily's bedroom, walks silently across the room, then stops and looks at her for 3 s:

(4) What is going on

Sim = Simon, father (gaze +, embodiments %); Emi = Emily, 19-year-old daughter (gaze *, embodiments •).

```
01          %                    (3.0)              %+   (2.0)   +
   sim:     %walks across Emi's bedroom to end of her bed%+gaze to Emi+
   sim:
02 Emi:→    +What.
   sim:     +to E --->>
03          (1.2)
04 Sim:     What (.) is going on.
05          (0.4)
06 Emi:     'Tcha mean what's going on.=
07 Sim:     %It's% ten to eleven, you gotta cab [waitin' outside,
            %brings L arm up and glances at watch%
08 Emi:                                         [Yeah becuz I'm la:te.
09          (0.8)
10 Emi:     Yeah.
11          (.)
12 Sim:     Yer ill.
13          (1.1)
14 Sim:     I'M ANSWERING THE DOO:R, (.) >I mean< WHO THE BLOODY HELL
15          D'YOU THINK YOU A:RE?
```

Simon's (delayed) response to Emily's "What" here delivers, in l.4, a complaint in the form of a question: "what is going on." This complaint itself gets repair initiated upon it by Emily and is subsequently elaborated on by Simon. Emily's defense (l.8), selectively addressing the fact in Simon's turn concerning the cab, but not the manifest complainable about the time, only prompts Simon to continue listing her offenses (l.12 and l.14) and then produce an assessment in the form of a rhetorical question (l.14-15), so escalating the conflict. In due course, this exchange ends with Emily storming out of the house, leaving her parents in visible distress.

Once again, then, a sustained gaze is responded to by "What"—a trajectory leading ultimately to conflict—but the response to it shows that it is treated distinctly in (2) and (4). What Simon's response to Emily in (4) shows us is that the sustained gaze he directs at her is in fact the complaints he delivers in l.7, 12, 14 and 15. In other words, he treats Emily's "What" as a go-ahead (Schegloff, 2007, p. 30) to his complaint. In (5) below, we see similarly a "What" being treated as a go-ahead (at l.2, and then subsequently at l.8) to a turn which is a pre-announcement ("Y'know w't I did las'night?", l.1):

(5) A terrible thing

Schegloff (1997), pp. 516–517; Hyl = Hyla; Nan = Nancy.

```
01 Hyl:     Y'know w't I did las'ni:[ght?
02 Nan:→                            [Wha:t,=
03 Hyl:     =Did a terrible thi::::[ng,
04 Nan:                            [You called Si:m,
05          (0.4)
06 Hyl:     No:,
07          (.)
08 Nan:→    What,
09          (.)
10 Hyl:     .t.hhhh[Well I hed-]
11 Nan:            [You called ]Richard,=
12 Hyl:     =(h)y(h)Yea(h)h en I h(h)ung up…
```

Schegloff further notes that:

> Forms of turn-constructional unit that can be used to initiate repair on another's prior turn can also be used as types of response in what we have come to call "presequences" of various types. The so-called "generic pre-sequence," which serves advance notice of some upcoming "business" without marking what that is, is the summons/answer sequence (1997, pp. 513–514).

In (6) below, we see one such exemplar. Amidst other ongoing activities, Fred is summoned by his mother (l.2), to which he responds with the aligning go-ahead "What" (l.3):

(6) Salami (Schegloff, 1997, p. 514) Nao = Naomi; Fre = Fred; Ann = Anne, mother.

```
01  Nao:       .hh if you [do this]one [you wi   ][n,
02  Ann:                                 [Fredeluh,][      ]
03  Fre:→                                          [Wha:t,]
04  Ann:       Wouldju wan[na give yer dog a li'l salami, ]
05  Nao:                  [En if you do only these, you lo]se.
06  Fre:       Oh (  )
```

There is thus independent evidence that "What" produced by Tom in (2), notwithstanding Emily's disingenuous treatment of it as an other-initiated repair, is clearly a "go-ahead" to the prefacing action constituted by the sustained gaze—as indeed is that produced by Emily herself in (4). In structural terms, then, the look is a so-called "pre," serving, in Schegloff's words, as "advance notice of some upcoming business" (Schegloff, 1997, p. 514). However, we have already seen enough evidence to suggest, *contra* Schegloff, that unlike the summons, which may be heard as equivocal with respect to what it prefaces, as a "pre" the look is not necessarily treated as a neutral action. Shay's response to Sunny's fixed gaze (for all that it is not directed at her) and the responses in extracts (2) and (4) suggest a technical preference for understanding the look to adumbrate some kind of challenging or problematic action for the recipient—an understanding that is indeed borne out by the interactions that follow.

5 Fixed gazes as adumbrating a problem

The origins of the fixed gaze as adumbrating a problem can be identified in the turn-taking system and specifically the practices for selecting the next speaker. Both Goodwin (1980) and Lerner (2003) discuss gaze as one resource for selecting next speaker; as Lerner notes: "It is common for speakers to look at or look for an addressed recipient as they begin to speak, and for the onset of a speaking turn to occasion a reciprocal gaze by coparticipants to determine if they (alone) are being addressed" (2003, p. 180; see also Goffman, 1963 and Cary, 1978 for observations on gaze as initiating encounters). The common retort "Don't look at me" responds to the implication, carried by a gaze, that the gaze producer is about to prevail upon the gaze recipient to act in some way.[7] Such an implication aligns with the norms of progressivity:

> '…Moving from some element to a hearably-next-one with nothing intervening is the embodiment of, and the measure of, progressivity…' (Schegloff, 2007, p. 15)

So, when a gaze producer does not produce a "hearably-next" action, stalling progressivity, and, in its stead, fixing the gaze on a recipient with an impassive expression, the gaze becomes, at the very least, accountable. In the following, this accountability rises to the surface of the talk following a fixed gaze of 0.8 s (l.11):

(7) Panera Drive-Thru

(25_FCSp14_AT_: 26:30 Family in the kitchen 1 Mandelbaum, Rutgers)

Dau = Daughter (* eye gaze*, $embodiments$); Mom, (+eye gaze+, &embodiments&).

The family is cooking together, with Mom and Daughter making up a packet of taco sauce. Dau is at the hob stirring the sauce while Mom takes the packets out of the cupboard to hand to Dau.

```
01 Dau:   *I mean they had a Panera Drive-Thru:.*
          *to Mom*
02        &(0.8)&
     mom: &Turns towards Dau with three packets, holds up to Dau&
03 Mom:   &°Can you open these? (----)°
     mom: &Extends packets towards Dau; one drops, takes it and
          places it in front of Dau
04        (0.4)&
05 Mom:   All five of 'em.
06  Dau:  Panera Drive-Thru.
07        $(0.4)$
     dau: $Shakes a packet as Mom studies the back of another$
08 Mom:   +That's fine.
          +reads back of packet she's holding -->
09        (1.5)+   +(0.8)
            --> +   +Looks up at Dau and holds gaze-->l.11
10 Dau:→ #$Hm?
           $chin thrust to Mom; mutual gaze in hold-->
    Fig:  #4
11        &(0.5)&$+
     mom: &Shakes head, then returns gaze to packet&+
     dau:     --->$
12 Dau:→$Why'y'lookin'me like that.$
          $furrows brow, gaze still at Mom$
13 Mom:   &Jis'* figuring this out.&
          &gazing at packet, RH goes to scratch head&
     dau: ---> *
14        (0.4)
     mom: $&turns away from Dau towards cupboard&$
     dau: $tears packet open & leans to shake contents in pan$
```

7 This is embodied in reports of episodes where a mutual gaze prefaces particular actions—in the following newspaper headline, an astonished response to an exceptional occurrence: "'We just looked at each other and started laughing': Raynes Park mum defies odds of 'one in 200 million' to give birth to identical triplets" *Your Local Guardian*, 20th September 2016.

FIGURE 4
Mom's gaze to daughter and daughter's "Hm?," extract 7, l.10.

Just as in extract (2) and indeed (4), the response to the look, here "Hm?," deploys a form commonly used for other-initiated repair, but which is here a go-ahead (see Figure 4).

But before Mom responds, Daughter follows up with a challenging question: "Why'y'lookin'me like that" (l.12), putting on record the accountability of the fixed gaze. Mom's response, unlike those in (2) and (4), defuses the potential hostility in its account: "Jis' figuring this out" (l.13).

In some cases, the go-ahead is omitted altogether, and the gaze producer is called directly to account. In the following, Jane enters the bedroom in the wake of a row with Emily and stands still for 5.8 s, looking at Simon as he finishes his phone call. He turns around to see her gazing fixedly at him:

(8) What you looking at me like that for

Sim = Simon (*eye gaze*, +embodiments+); Jan = Jane (%eye gaze%, $embodiments$).

```
01          $%    (5.8)   %$
    jan:    %    to Sim   %
            $standing still$
02  Sim:    (on phone) Alright, bye!
03          +(4.4)+
            +hangs up phone, turns round to see Jan gazing at him, stops+
04  Sim: →  *What you l(h)ookin' at me like that for.=
            * to Jan -->>
05  Jan:    =Well because you don't back me ↑U::P!!
06          (0.3)
07  Sim:    What is it you want me to ↑DO::::?
08  Jan:    ↑W'l (.) instead'v goin', ooh,you know, d- you don'wanna
09          start that now
10          (0.8)
11  Sim:    W'l you've walked through the doo:r, an' you've gone straight
12          for'er
```

Simon's "What you l(h)ookin' at me like that for" (l.4)—itself a hearably upgraded format from the responsive "why" used in (7)—similarly calls the gaze-producer to account, "like that" explicitly pointing to a way of looking that is being challenged. Here, its infiltrated laugh token indicates the delicacy of the challenge (see, e.g., Clift, 2012 on laughter marking delicacy) and gets an immediate, latched response from Jane: a complaint with some articulatory force and prosodic animation "Well because you do not back me U::P!!" (l.5). Simon's response—another challenging question—almost exactly mimics Jane's prosody and articulation, and thus escalates the conflict.

A more escalated exchange still, following a fixed gaze, can be seen in the following. As in the previous extract, the gaze is treated as accountable—and, as in the previous extract, the format "what…for" is used. However, the gaze itself is produced in an environment that is already fissile. Emily has been summoned to the living room by her parents to address her recent wayward behavior. She arrives and crosses the room, watched by her parents, bounces down on the sofa, clears her throat, sniffs (all actions that might be hearable as preparing to speak), but then fiddles with a blanket, and then lifts her gaze to fix it on her mother:

(9) Filthy Looks

Jan = Jane; Emi = Emily (gaze *, embodiments •); Sim = Simon (gaze +, embodiments %).

```
01          •    (0.4)  • *(0.8)** (0.6)    * •    (0.8) •
    emi:    •raises head• *to Jan**gaze down* •jerks head•
02          * (1.8)
    emi:    *#to Sim -->?
    Fig:     #5
03          %     (1.7)    %
    sim:    %eyes close for 1 second, then rapid blinking -->
04  Jan: →  What's %the %filthy % %looks for.%
    sim:        -->%    %eyeroll% %blinking
05  Emi:    *Becu:z! I've sai- I've trie- I've said to Dad I don't
            *to J -->>
06  Emi:    wanna have this convers↑ation.
```

From the time Emily sits down to Jane's response is a whole 12 s, and her gaze from when she lifts her head to gaze at her parents, over

FIGURE 5
Emily's gaze at extract 9, lls.1–3. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

6 s (see Figure 5)—all while her parents sit silently watching her. Jane's response to this gaze is, as in (8), a challenging question—but, while in (8) Simon's assessment of Jane's look was not made lexically explicit, here Jane's assessment of the look is made plain. The reference to "filthy looks" constitutes a double upgrade of the familiar phrase "dirty look": lexically, from "dirty" to "filthy," and morphologically, from one such "look" to several. Emily's response does not challenge either the question or the assessment it embodies. Her emphatic production of "Becu:z!" as its own TCU initially resists producing a further account, but an account does then follow, albeit with some dysfluency attending its launch—and one that grounds the hostility of the look in her reluctance to engage—one that is reported, in an epistemic upgrade (Clift, 2007), as having been registered with her father earlier: "I've said to Dad I do not wanna have this conversation" (ll.5–6). What subsequently follows is a highly antagonistic exchange.

We have thus seen a number of instances in which a fixed gaze is taken to be adumbrating some kind of problem or challenge. In the most benign context (7), we see a gaze being called to account; the subsequent response works both to account for the look and to reassure, and conflict is thereby averted. In (8) and (9), the fixed gaze initiates a conflict sequence, just as had (2) and (4). However, as we have seen, these sequences themselves all follow from prior episodes of conflict – either immediately prior, as in (4), (8), and (9), or, as Emily makes clear in (2), some time previously: "Tom did not… need to…say anything to me earlier" (lls. 14–15). So, in these cases, the fixed gazes can be seen as the initiating action in a potential renewal of hostilities.

This is not to suggest, of course, that all fixed gazes initiate conflict sequences. Extract (7), where there had been no prior antagonism between the parties, shows a gaze addressed and accounted for unproblematically. It is clear, then, that in some contexts a fixed gaze may be managed to more peaceable outcomes.

6 Defusing hostility

While the majority of instances in our corpus, as represented in the cases mentioned above, showed gazes initiating a trajectory of conflict, we only found two cases involving fixed gazes where such conflict was averted, one by the gaze-producer, and one by the gaze-recipient; in different ways, they throw into relief the characteristics of the conflict-initiating gaze. The first is a case in which the gaze is accompanied by (lighthearted) verbal indications of trouble, and so in that respect distinct from the cases examined so far. Here, Simon sits having breakfast with Emily and Tom. He gazes over the table at Emily, who spent the previous night at a nightclub, "Icon," and whose eye-makeup is visibly smeared:

(10) Black eyes F1:2 3:52–4:06

Sim = Simon (*gaze* + embodiments+); Emi = Emily ($gaze$ %embodiments%).

```
01              *   (3.0)  *
        sim:    *  to Emi  *
02      Sim:     $*£Ooh dear-£*
        emi:     $  to Simon  --> l.5
03               (0.2)
04      Sim:     huh huʰ huʰ
05               #(0.2)
        Fig:     #6
06      Emi: →   What.$
                   -->$
07      Sim:     $heh £G(h)ot ra(h)ther %bla(h)ck e(h)yes l(h)o:ve. £
        emi:     $looks down at sandwich she is holding$
        emi:                              %raises sandwich to mouth%
08               (0.2)
09      Emi:     $°Oh well.°$
                 $at sandwich$
10      Sim:     .hhhh >where d'you go last ni:ght.<
11               (0.4)
11      Emi:     Icon.
```

Simon's fixed gaze is thus accompanied by his verbal trouble-alert, "Ooh dear" and then a laugh, as he looks at Emily. Her go-ahead "What" (l.6) in response is produced as she in turn gazes at Simon with visible wariness (Figure 6); notwithstanding Simon's jocular overtures, Emily declines to join him in his laughter, hearable as it is as a potential tease, laughing, not with, but at her.

Simon's response to the go-ahead, "£G(h)ot r(h)ather bla(h)ck e(h)yes l(h)ove£" (l.7)—a negative assessment accounting for his prior trouble-alert and laughter, is itself laughter-filtrated. At this, Emily, looking at her bacon sandwich, lifts to her mouth, saying "Oh well"—a display of resignation that registers Simon's assessment without taking issue with it. Simon's highly redressive action in producing his gaze underscores the fact, demonstrated by extracts (2), (4), (7), (8), and (9), that it is the fixed and silent gaze that is taken to be potentially hostile.

The other instance where a gaze did not escalate into conflict is one where the recipient figures what the gaze is adumbrating, and takes measures to defuse it. The context is a highly affiliative one—a Valentine's meal, prepared by Simon for Jane. As she prepares to eat, she compliments the paella in front of her (l.1), but as she raises the

pepper mill, visibly about to grind pepper on her food, she sees Simon gazing at her. She lifts the pepper mill, points at, and, with a straight face (Figure 7), produces a defense to what she thereby implies is about to be his complaint: "Tasted it first":

(11) Tasted it first

Jan = Jane (*gaze* + embodiment+) Sim = Simon (%gaze% &embodiment&).

```
01  Jan: +°This is ni:ce.°+
         +transfers lemon wedge from R to LH+
02              +(0.2)+                    +(2.0)
         +pulls plate towards her+ +picks up pepper mill and keeps
         in #hold -->>
    Fig:    #7
03  Jan: I've tasted it fi:rst¿
04       %(1.0)%    %&(0.2)
     sim:     %at J%    %to middle distance-->>
     sim:                &#smiles -->>
     Fig:                    #8
```

FIGURE 6
Emily's response to Simon's gaze, just before "What" (l.5), extract 10, l.4. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

FIGURE 7
Jane points to pepper mill, extract 11, l.2. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

Simon, in response, produces a wry smile (Figure 8).

While reminiscent of Kidwell's (2005) sanctionable gazes by caregivers, the distinction is that here the adult interactants share a history, and one that is likely to have been drawn on here by Jane. "I've tasted it first" (l.3) invokes a history of exchanges where Simon has cause to complain that his cooking is being seasoned before having been tasted. Jane's response thus pre-empts this adumbrated complaint, her retort deftly defusing the nascent hostility.

FIGURE 8
Simon's response to Jane, extract 11, l.4. © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

In this connection, it is perhaps unsurprising that the dining table should figure in a number of the instances [(1), (2), (10), and (11)] here. As a site for a (standardly) daily gathering, opportunities for monitoring—and so potentially sanctioning—others' behavior are both ample and recurrent, facilitated by the positions of the participants in Kendon's F-formation, either in so-called vis-à-vis or L-arrangements (Kendon, 1990, p. 209).

Although these last two instances show fixed eye gazes initiating a trajectory that does not lead to conflict, the fact that, in both cases, redressive action to avert conflict needs to be taken—in (10) by the gaze-producer, and in (11) by its recipient—itself constitutes evidence of the conflict-initiating potential of such gazes.

7 Conclusion

The potential hostility attached to a studied, silent gaze in public spaces among strangers was found to be overwhelmingly replicated in our data, recorded in domestic settings among intimates. Structurally, this fixed gaze is a so-called "pre" action, designed to adumbrate some business—and that, overwhelmingly, the nature of that business is taken to be challenging in some form. In such cases, the designed eye gaze initiates an interactional trajectory toward conflict. A few cases we have shown constitute exceptions to this usual trajectory from studied eye gaze to conflict. The first instance, in (1), showed a fixed gaze mistakenly taken to be hostile. Extract (7) showed a gaze being treated with wariness, and being accounted for by its producer. In both, there was clearly an orientation to the projection of a potential negative stance, neither of which turned out to be so designed. Extract (10) was a gaze which was, like that in (7), treated warily but mitigated with verbal resources and laughter and so conflict was headed off. Moreover, extract (11) also shows conflict averted by the gaze-recipient with an account for the action taken to be the target of the projected complaint. There were no instances in the data of a fixed and silent gaze designed to be preliminary to an unequivocally affiliative action. The single instance of the latter identified in the course of this study was one, not recorded, but reported on social media—and not between intimates, but strangers. In the wake of Elon Musk's reported

FIGURE 9
Twitter, 20th December 2017, Transit Line Connection - What?!?" via Twitter @KenOhrn.

complaints in 2017 about traveling by public transport, several responded on what was then Twitter with stories about the great things that had happened when they had taken busses and trains, such as meeting their spouses or best friends. The following relates one such encounter, initiated with a studied gaze—the reportability of an affiliative moment on public transport underscored here by the wordlessness[8] of the exchange (Figure 9).

As reported here, the studied gaze from a stranger prompts a (from the lack of quotation marks, silent) go-ahead "What"—the punctuation suggesting not so much aggression but bemusement—the response to which is an embodied offer, then graciously declined. That this reported instance was the only one identified in which a fixed gaze is designed to adumbrate a wholly positive action further attests to its relative rarity, at least in the data of English interaction.[9]

This empirical skewing—moreover, on the data of intimates, not strangers—thus suggests an overwhelming preference for taking the fixed gaze to be projecting an upcoming negatively polarized action. This observation dovetails with the preference for progressivity and the corresponding association of bodily holds with at-that-moment unresolved issues or problems in the talk. The recipient of such a gaze, as we have seen, can then produce a responsive go-ahead (such as "What" or "Hm?") which maintains a relatively neutral stance with respect to what is likely to come, or an immediately more confrontational one, demanding an account for the characteristics of the look (e.g., "Why you looking at me like that for"/"What's the filthy looks for"). Such actions may be initiating a verbal course of action, but of course have their source prior to the talk, in the fixed gaze. In this respect, the sequences launched by the gaze are structurally what Schegloff (2007) calls "retro-sequences," invoking a source-outcome relationship, in which, he notes:

> …the first recognizable sign that such a sequence is in progress generally displays that there was "a source" for it in what preceded, and often locates what that source was. But note that the source engendered nothing observable—indeed, was not recognizable as "a source"—until the later utterance/action, billing itself as an "outcome," retroactively marks it as such. Their "firstness" follows their outcome, though their occurrence preceded it. These are sequences *launched* from their *second* position (2007, p. 217).

Locating the source involves monitoring what in the environment might be called to account (hence, as noted earlier, the dining table as a particularly rich site for possible candidates, where one might take another to be judging them for, e.g., slurping food or licking a knife); Shay's query to Sunny, "Are you angry with me over something I've said" in (1) exactly captures this uncertainty. In this respect, Eckert's concept of an "indexical field" as a set of possible interpretations that undergo indexical specification *in situ* (Eckert, 2008, p. 454; see also Heritage, 2016, p. 209) is a useful one. A studied gaze evokes an indexical field of possibilities, narrowed down by both compositional features (such as facial expression) and sequential context. The resonance of this with Garfinkel's (1967, p. 34) characterization of the indexical and reflexive properties of language and action is clear.

The instances we have examined, in the data of mundane domestic interaction, thus show very clearly how one practice—the fixed gaze—can in fact initiate an interactional trajectory that may be increasingly conflictual. Schegloff, contemplating the extreme outcomes to which such trajectories may ultimately lead, observes that understanding how such horrors arise is the first step in attempting to address them:

> Rape, abuse, battering, etc., do not exist in some other world, or in some special sector of this world. They are intricated into the texture of everyday life for those who live with them. How else are we to understand their explosive emergence where they happen if not by examining ordinary interaction with tools appropriate to it, and seeing how they can lead to such outcomes…how else—when confronted by the record of singular episodes—are we to understand their genesis and course, how else try to understand what unwilling participants can do to manage that course to safer outcomes, how else try to understand how others might intervene to detoxify those settings? (Schegloff 1999, pp. 561–562).

In examining the sequential implications of a particular kind of look, it has thus been possible, as a first step in such an endeavor, to illuminate the normative assumptions of accountability and progressivity that underlie the resources we deploy in pursuing courses of action.

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.

8 The silent nature of the exchange is not the focus here, but it is possible that this may be fitted to the ambient noise on public transport.

9 A referee for this study contests the relative rarity of such positive actions compared to negative ones as adumbrated by a fixed gaze. This may be attributable to cultural variation. In my experience (and thus anecdotally), I have observed greater tolerance for the fixed gaze—and thus greater likelihood that it adumbrates some positive action, such as a compliment—in some cultures than others; hence, the caveat that the findings are observable only for English.

Ethics statement

The studies involving humans were approved by University of Essex Ethics Review System. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements. 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

RC: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft.

Funding

The author declares that financial support was received for the research, authorship, and/or publication of this article. This article was written while I was in receipt of a Distinguished Scholar Award from the Harry Frank Guggenheim Foundation for the project "Behind Closed Doors: Trajectories to Violence in Intimate Interaction," and I gratefully acknowledge the support of the Foundation, and for the Arts and Humanities Impact Acceleration Fund at the University of Essex in undertaking this research.

Acknowledgments

All photos and video clips from Figures 1–3, 5–8 © Dragonfly Film and TV Productions Ltd. Reproduced with thanks to Dragonfly Productions.

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

Argyle, M., and Dean, J. (1965). Eye-contact, distance and affiliation. *Sociometry* 28, 289–304. doi: 10.2307/2786027

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Soc. Psychol.* 41, 269–271. doi: 10.2307/3033565

Clift, R. (2007). "Getting there first: the non-narrative use of reported speech in interaction" in Reporting talk: Reported speech in interaction. eds. E. Holt and R. Clift (Cambridge: Cambridge University Press), 120–149.

Clift, R. (2012). Identifying Action: Laughter in Non-Humorous Reported Speech. *J. Pragmat.* 44, 1303–1312.

Clift, R. (2020). Stability and visibility in embodiment: the 'palm up' in interaction. *J. Pragmat.* 169, 190–205. doi: 10.1016/j.pragma.2020.09.005

Clift, R. (2024). "Beyond the lamp-post" in Handbook of research methods in conversation analysis. eds. J. D. Robinson, R. Clift, K. H. Kendrick and C. W. Raymond (Cambridge: Cambridge University Press).

Clift, R., Drew, P., and Local, J. (2013). "Why that, now? Position and composition in interaction (or, don't leave out the position in composition)" in Language, music and interaction. eds. M. Orwin, C. Howes and R. Kempson (London: College Productions).

Clift, R., and Rossi, G. (2023). Eyebrow raises in the transition space. *Soc. Interact.* 6:3. doi: 10.7146/si.v6i3.142897

Darwin, C. (1872). The expression of the emotions in man and animals. London: J. Murray.

Eckert, P. (2008). Variation and the indexical field. *J. Socioling.* 12, 453–476. doi: 10.1111/j.1467-9841.2008.00374.x

Eisenbarth, H., and Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 860–865. doi: 10.1037/a0022758

Floyd, S., Manrique, E., Rossi, G., and Torreira, F. (2016). Timing of visual bodily behavior in repair sequences: evidence from three languages. *Discourse Process.* 53, 175–204. doi: 10.1080/0163853X.2014.992680

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Gibson, J. J., and Pick, A. D. (1963). Perception of another person's looking behavior. *Am. J. Psychol.* 76, 386–394. doi: 10.2307/1419779

Goffman, E. (1963). Behavior in public places. New York: Free Press.

Goffman, E. (1971). Relations in public. New York: Basic Books.

Goodwin, C. (1979). "The interactive construction of a sentence in natural conversation" in Everyday language: Studies in ethnomethodology. ed. G. Psathas (New York: Irvington Publishers).

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn- beginning. *Sociol. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Goodwin, C. (1981). Conversational organization: Interaction between speakers and hearers. New York: Academic Press.

Green, M., Williams, L., and Davidson, D. (2003). In the face of danger: specific viewing strategies for facial expressions of threat? *Cognit. Emot.* 17, 779–786. doi: 10.1080/02699930302282

Griffin, J. H. (1961). Black Like Me. Boston: Houghton-Mifflin.

Groeber, S., and Pochon-Berger, E. (2013). Turns and turn-taking in sign language interaction: a study of turn-final holds. *J. Pragmat.* 65, 121–136. doi: 10.1016/j.pragma.2013.08.012

Heritage, J. (2011). "Territories of knowledge, territories of experience: empathic moments in interaction". In *The Morality of Knowledge in Conversation*. eds. T. Stivers, L. Mondada and J. Steensig. (Cambridge: Cambridge University Press).

Heritage, J. (2016). On the diversity of 'changes of state' and their indices. *J. Pragmat.* 104, 207–210. doi: 10.1016/j.pragma.2016.09.007

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychon. Bull. Rev.* 27, 856–881. doi: 10.3758/s13423-020-01715-w

Hietanen, J. K. (2018). Affective eye contact: an integrative review. *Front. Psychol.* 9:1587. doi: 10.3389/fpsyg.2018.01587

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendon, A. (1990). Conducting interaction: Patterns of behavior in focused encounters. Cambridge: Cambridge University Press.

Kidwell, M. (2005). Gaze as social control: how very young children differentiate "the look" from a "mere look" by their adult caregivers. *Res. Lang. Soc. Interact.* 38, 417–449. doi: 10.1207/s15327973rlsi3804_2

Lerner, G. H. (2003). Selecting Next Speaker: The Context-Sensitive Operation of a Context-free Organization. *Lang. Soc.* 32, 177–201.

Li, X. (2014). Multimodality, interaction and turn-taking in mandarin conversation. Amsterdam: John Benjamins.

Manrique, E. (2016). Other-initiation of repair system in argentine sign language (LSA). *Open Linguist.* 2, 1–34. doi: 10.1515/opli-2016-0001

Rossano, F. (2012). Gaze behavior in face-to-face interaction. Unpublished PhD Dissertation,. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

Schegloff, E. A. (1997). On practices and actions: boundary cases of other-initiated repair. *Discourse Process.* 23, 499–545. doi: 10.1080/01638539709545001

Schegloff, E. A. (1999). 'Schegloff's texts' as 'Billig's data': a critical reply. *Discourse Soc.* 10, 558–572. doi: 10.1177/0957926599010004006

Schegloff, E. A. (2007). Sequence Organization in Interaction. Cambridge: CUP.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The Preference for Self- Correction in the Organization of Repair in Conversation. *Lang.* 53, 361–382.

Sikveland, R., and Ogden, R. (2012). Holding gestures across turns: moments to generate shared understanding. *Gesture* 12, 166–199. doi: 10.1075/gest.12.2.03sik

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 November 2024
DOI 10.3389/fpsyg.2024.1467185

Check for updates

OPEN ACCESS

EDITED BY
Geert Brône,
KU Leuven, Belgium

REVIEWED BY
Anna Inbar,
University of Haifa, Israel
Kaoru Amino,
Keimyung University, Republic of Korea

*CORRESPONDENCE
Cornelia Müller
✉ cmueller@europa-uni.de

RECEIVED 19 July 2024
ACCEPTED 25 September 2024
PUBLISHED 27 November 2024

CITATION
Müller C, Karpiński M, Kindler-Mathôt C, Klessa K, Jarmołowicz-Nowikow E, Junge JK, Papadopoulou K and Sawicka-Stępińska B (2024) Affectivity as stance: multimodal stance-taking in audiovisual documentations of Polish and German parliamentary debates. *Front. Psychol.* 15:1467185. doi: 10.3389/fpsyg.2024.1467185

COPYRIGHT
© 2024 Müller, Karpiński, Kindler-Mathôt, Klessa, Jarmołowicz-Nowikow, Junge, Papadopoulou and Sawicka-Stępińska. 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.

Affectivity as stance: multimodal stance-taking in audiovisual documentations of Polish and German parliamentary debates

Cornelia Müller[1]*, Maciej Karpiński[2], Clara Kindler-Mathôt[1], Katarzyna Klessa[2], Ewa Jarmołowicz-Nowikow[2], Jana Katharina Junge[1], Katerina Papadopoulou[1] and Brygida Sawicka-Stępińska[2]

[1]Faculty of Social and Cultural Sciences, European University Viadrina, Frankfurt (Oder), Germany, [2]Faculty of Modern Languages and Literature, Adam Mickiewicz University, Poznań, Poland

This paper presents a media-aesthetic framework to study affectivity as a stance. This framework opens up a new perspective on multimodal affective stance-taking in the context of specific media ecologies. It exemplifies this new approach with case studies of the official audiovisual documentation of political debates in the German Bundestag and the Polish Sejm. This new approach addresses the intertwining of audiovisual multimodality with verbo-gestural expressivity (or the multimodality of speaking). Adopting a phenomenological position, we are interested in how the orchestration of the debates as audiovisual events moves the spectators. The concepts 'Expressive Movement' and 'Dynamic Forms of Vitality' serve as theoretical and methodological references to capture the affective dynamics of audiovisual debates and how these audiovisual images modulate the perceptions and experiences of the spectators. To illustrate and substantiate this approach for linguistic and media analyses of affective stance-taking, the paper outlines basic assumptions and methods. It offers two exemplary case studies from German and Polish parliamentary debates. It is concluded that bringing together media-aesthetic with linguistic analyses of multimodal communication and interaction provides not only a valid starting point for future research of multimodal stance-taking in different media ecologies but also allows researchers to address how and why spectators of audiovisual media performances are moved affectively.

KEYWORDS

multimodality, gesture, media analysis, political communication, discourse analysis, interaffectivity, multimodal discourse analysis, prosody analysis

1 Introduction

It is common practice in democratic societies to document the work of the parliament in an archive. Public debates of bills and budgets are one of the central activities of a parliament. Such archives function as "memories" of the parliament, but they also invite the public to participate in the debates of their elected representatives. How such political debates 'move' the spectators is the subject of a German–Polish research project, which explores multimodal affective stance-taking in the German Bundestag and the Polish Sejm. The project investigates audiovisual materials supplied by the respective archives of the budget and financial debates in the German and the Polish parliaments. Analyzing multimodal stance-taking in the context of audiovisual documentations of parliamentary debates requires a reflection on both the

mediatized character of the material *and* the multimodal performance of the speakers and their engagement with the audience in parliament. Put differently, analyses face an intertwining of audiovisual multimodality and verbo-gestural expressivity (or what we consider: the multimodality of speaking). This has important consequences for the research into stance-taking.

Despite a rich and diverse literature on stance and stance-taking (Avdan, 2017; Biber and Finegan, 1989; Dancygier, 2012; Du Bois, 2004, 2007; Englebretson, 2007; Feyaerts et al., 2017; Haddington, 2006), the role of gesture, posture, and prosody, as aspects of this communicative activity, remain only scarcely studied. Although Du Bois' paper on the stance triangle mentions gesture several times as one of the expressive forms that contribute to stance-taking (Du Bois, 2007, pp. 169, 171) and suggests speaking of stance-taking as an activity rather than of stance as a phenomenon of lexical semantics and text linguistics (Biber and Finegan, 1989), his insight did not foster a systematic inclusion of gesture, speech, and body movement in the study of stance-taking. Some of the few exceptions come from anthropology and conversation analysis (Goodwin et al., 2012; Ochs and Schieffelin, 1989) and gesture studies (Bressem and Müller, 2017; Ladewig, 2014, 2024; Müller and Speckmann, 2002), including shrugs and head tilts (Debras and Cienki, 2012). The role of prosody in stance-taking has been directly explored to a limited extent (Freeman, 2015a, 2015b). More recently, and in a plea to integrate conversation analytic and interactional perspectives on stance with the usage-based approach of cognitive linguistics, Feyaerts et al. (2017) lay out the significance of multimodal aspects of stance-taking specifically with regard to alignment. Cognitive linguist Dancygier (2012) introduces the notion of 'stance-stacking' to indicate that expressions of stance form constructions and often do not come alone, but build clusters. Dancygier's study highlights the semantic complexity of stance and shows how language use in the media is permeated by visual modes of expression to a degree that they develop into multimodal constructions. Dancygier and Vandelanotte (2017) exemplify this with constructional analyses of internet memes. Taking an expressive stance is seen as *the* motivation behind making a meme.

The present paper presents a media-aesthetic framework to study affective multimodal stance-taking in audiovisual media. It offers a new view on embodied interaction in multimodal stance-taking, in both face-to-face interaction and when spectators engage with televised face-to-face interaction. This position draws on earlier study by Ochs and Schieffelin (1989) and Goodwin et al. (2012) and on media-aesthetic (Kappelhoff, 2015) and neo-phenomenological film theory (Sobchack, 1992). Ochs and Schieffelin (1989, p. 7) suggest that linguistic and discourse features "provide an affective frame for propositions encoded" and highlight the relevance of "gestural cues to provide interlocutors with critical information on which to base subsequent social actions." Following up on this study, Goodwin et al. (2012) regard affective stance as "situated practice entailed in a speaker's performance" and achieved "through intonation, gesture, and body posture" (Goodwin et al., 2012, p. 16).

The position advocated in this paper takes recourse to the idea of multimodal communication as a dynamic, intercorporeal, and hence *interaffective* process (Müller and Kappelhoff, 2018, chapters 1–5). Conceiving *affectivity as stance* thus addresses a basic form of human understanding that is grounded in an intercorporeal exchange of perceiving, experiencing, feeling. In such a phenomenological view, perceiving is experiencing the other as a *Moving Other*. Research on social interaction has shown that when people speak with one another, this being together is established through subtle forms of bodily synchronization (Kendon, 1972), it is maintained by mutual alignment (Feyaerts et al., 2017) and rhythm (Breyer et al., 2017). Phenomenology regards this intercorporeal being together as an *interaffective* dimension of experience (Fuchs, 2017). In order to grasp this interaffective dimension of experience in contexts of televised multimodal performances—as both a theoretical concept and empirical methodology, we work with the notions of 'Expressive Movement' (EM) and 'Forms of Vitality'.

For the study of affective stance, we extend previous research on cinematic EMs as interaffective embodied experiences that ground metaphorical meaning-making in both face-to-face interaction and the process of viewing moving images (Horst et al., 2014; Kappelhoff and Müller, 2011; Müller and Kappelhoff, 2018). In our study, stance-taking is conceived as a public activity in which the complexities of multimodal stance expressions form multidimensional experiential gestalts that are permeated by affect—theoretically and analytically graspable as EM (Müller and Kappelhoff, 2018; chapters 2, 3, 6, 8).

To investigate the affectivity of multimodal stance-taking in audiovisual media we have conducted a research project on political debates in the German and the Polish parliaments, respectively. The paper presents insights and the first results from this larger undertaking. It begins with a brief outline of the basic assumptions of a media-aesthetic approach to affectivity as stance in section (2); section (3) introduces data and methods; (4) presents a comparative case study of multimodal affective stance-taking in parliamentary speeches; (5) offers a summary and a discussion; and (6) concludes the paper.

Why investigate affective stance-taking in parliamentary debates and speeches? Clearly, parliamentary speeches are meant not only to provide factual information but also to impress and convince addressees. Therefore, they are a rich resource for studying affective stance-taking. As official documentations of political debate, they are meant to be as 'neutral' or 'objective' as possible, so we would expect the audiovisual staging to be rather neutral and not to play a prominent role in the affective perception of the televised embodied performances. We will see whether this is in fact the case.

2 A media-aesthetic approach to affectivity as stance

This chapter draws on earlier study on Cinematic Metaphor (Kappelhoff and Müller, 2011; Müller and Kappelhoff, 2018). Against the backdrop of this transdisciplinary approach, the present paper brings together film theory and linguistics to analyze televised forms of face-to-face interaction. The theoretical positions outlined in this section sketch a media-aesthetic framework for the multimodal analysis of affective stance in audiovisual documentations of political debates. Starting from the expressive character of multimodal communication, these basic theoretical assumptions lay the ground for media-aesthetic and linguistic methods for researching multimodal affective stance-taking. These theoretical points are particularly relevant here: affectivity as a dynamic form of vitality (2.1), Expressive movement as multidimensional experiential gestalt (2.2), and affective stance as a media phenomenon (2.3).

2.1 Affectivity as a dynamic form of vitality

In his study on the interpersonal world of the infant, Daniel Stern describes the attunement of parent and newborn child as a dance (Stern, 1985). Dancing together means moving together. It involves feeling the movement dynamics of the other person in one's own body. In his later study, Stern extends his observations from early child development to the "experience of vitality" as a dynamic form that "permeates daily life, psychology, psychotherapy, and the arts" (Stern, 2010, p. 3). Stern argues that "We naturally experience people in terms of their vitality" and that also the "Time-based arts, namely music, dance, theater, and cinema […], move us by the expressions of vitality that resonate in us." (Stern, 2010, p. 4). Drawing on Stern's approach, we consider affectivity as a *quality of movement* where 'movement' includes body movement, speech movement, and the movement of audiovisual images. Rather than conceiving of affect as an expression of some inner state discernible as facial expression (Ekman, 1972, 1992), affectivity is described as a form of experienced vitality, a dynamically *unfolding temporal contour*. Stern describes them as a crucial dimension of being:

> They are the felt experience of force – in movement – with a temporal contour, and a sense of aliveness, of going somewhere. They do not belong to any particular content. They are more form than content. They concern the "How," the manner, and the style, not the "What" or the "Why" (Stern, 2010, p. 7).

Stern directs our attention to a group of words that describe such forms of vitality as felt experiences, exploding, pulsing, or fading describes forms of vitality (Figure 1). An exploding form of vitality shows a steeply rising curve, a pulsing one, and a series of small curves on a steady horizontal level; a fading vitality is a continuously declining contour. Stern illustrates the "felt experience of force in movement as a temporal contour" as graphs of time and intensity (Figure 1).

Forms of vitality concern the dynamic expressive qualities of movement. They address how a movement is performed. They are bound to movement and time, and they are modality-independent. Here is where Stern connects bodily forms of vitality with time-based arts.

> The same is true for the time-based arts. The dynamic flow of music (sound in motion), dance, theater, and cinema sweeps us up at moments and then releases us, only to sweep us up again quickly just downstream (Stern, 2010, p. 6).

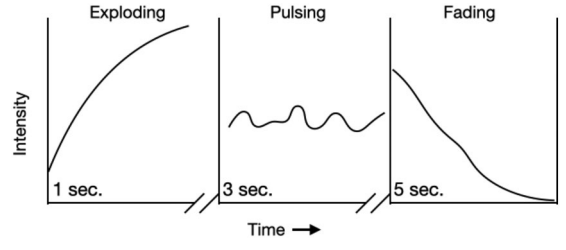


FIGURE 1
Stern's dynamic forms of vitality, three possible vitality forms (p8).

These commonalities between the time-based arts and body movements are essential for a media-aesthetic approach multimodal affective stance. They allow us to account for affectivity as a dynamic form of vitality that is expressed in the movement dynamics of both the delivered speech and the audiovisual documentation of it.

2.2 Expressive movement as multidimensional experiential gestalt

How precisely are these movement contours perceived? In order to understand this, we need to consider the perception of body movements further, and here is where the concept of Expressive Movement comes in (Kappelhoff, 2004, part 1; Kappelhoff, 2018b; Müller and Kappelhoff, 2018, chapter 9). It concerns a specific understanding of body movements. The concept is historically rooted in early 20th-century discussions in philosophical anthropology, psychology, and linguistics. Expressive Movement offers an alternative perspective on the idea of expression as an involuntary, outer index of a subject's inner disposition. Scholars like Bühler, Plessner, and Wundt shared the assumption of expressive behavior as an interplay of affective exchanges of intensity between living beings. In Plessner's philosophical anthropology, for example, Expressive Movements are not expressions of inner states, but a form of behavior that synchronizes an organism with its environment. Body movements evolve as a wholeness, a movement gestalt, be they affective or affectless:

> These wholenesses belong to the organism through its relation to the environment, […] As a result, the movement shapes are pictorial, even stretched out over a certain duration of time […] Grasping, fleeing, repelling, seeking, but also the "affectless" forms such as walking, flying, swimming […] represent such movement-images (Plessner, 1982 [1925], p. 78).

The Expressive Movements unfold as movement shapes, which become movement gestalts or movement images in the process of perception. Moreover, following Stern, body movements (including speaking) and audiovisual images are equally time-based forms of expressions, they not only unfold as a specific dynamic contour of movement but also they both become movement images in the perception of a co-participant or a spectator at a screen (see Kappelhoff's notion of "Cinematic Expressive Movement," Kappelhoff, 2004, part 1; Kappelhoff, 2008; Kappelhoff, 2018b; Müller and Kappelhoff, 2018). Expressive Movements are thus gestalts that emerge in a process of movement experience. They are orchestrated along multiple dimensions. For kinesic Expressive Movements, this includes speech, gesture, posture, gaze, and head movement; for cinematic Expressive Movements, this involves all the aspects of cinematic orchestration of audiovisual images, such as camera angle, camera movement, montage, sound, and mise-en-scène. The concept of Expressive Movement thus suggests a shared understanding of body movements and cinematic movement images.

The affective dimension of Expressive Movements *is* their quality of movement. This position aligns with Merleau-Ponty's phenomenology, where gestures of anger or threat do not express an inner emotion, but *are* anger or threat: "Faced with an angry or threatening gesture, I do not see anger or a threatening attitude as

a psychic fact hidden behind the gesture, I read anger in it. The gesture does not make me think of anger, it IS anger itself" (Merleau-Ponty, 2005 [1945], p. 184). Affectivity in this perspective is intercorporeal and interaffective, it is a form of embodied understanding. Expressive Movements are thus conceived as 'multidimensional experiential gestalts' that modulate and ground embodied processes of meaning-making affectively. Expressive Movements can be body movements or audiovisual movement images (Müller and Kappelhoff, 2018, chapters 1, 8, 9). In mediatized audiovisual documentations of political speeches, the dynamic unfolding and the internal orchestration of Expressive Movements frame the multimodality of speaking.

2.3 Affective stance in mediatized political debates

What are the consequences of such a position for an analysis of stance-taking in mediatized political debates? First, it underlines DuBois' understanding of stance as an inherently multimodal and undeniably complex action. Du Bois defines stance.

> "as a public act by a social actor, achieved dialogically through overt communicative means (language, gesture, and other symbolic forms), through which social actors simultaneously evaluate objects, position subjects (themselves and others), and align with other subjects, with respect to any salient dimension of value in the sociocultural field" (Du Bois, 2007, p. 169).

Second, it supports the idea of stance-taking as an activity rather than a phenomenon of lexical semantics and text linguistics (Biber and Finegan, 1989). It also underlines positions by Ochs and Schieffelin as well as by the Goodwins, who consider stance as a joint doing brought about by cooperating participants in an interaction (Goodwin, 2017; Goodwin et al., 2012; Ochs and Schieffelin, 1989).

But why consider the audiovisual orchestration of televised multimodal stance-taking? One could simply ignore the character of the audiovisual recording and directly analyze the verbo-gestural or multimodal performance of the speaker. Doing so would analyze what the analysts were able to see and leave aside movements not visible on the film image. Multimodal stance-taking would be reconstructed as a speaker's process only. However, aside from competing theoretical positions concerning the nature of speaking as an individual or an interactive process, this kind of analysis would give no clue as to how the political speeches are perceived by the audiences watching them on their screens. It would neglect the question of how audiovisual images affect their audiences so deeply—a question rendered even more acute since the popularity of TikTok videos. One answer to this has been formulated by media-aesthetic film theory: cinematic staging affects viewers bodily (Kappelhoff, 2004, 2018a, 2018b; Sobchack, 1992) (Figure 2).

Understanding film is not a cognitive puzzle of finding a storyline, it is experiencing cinematic movement images, and it is feeling the film. Neo-phenomenological film theorist Vivian Sobchack argues that it is spectators who lend the film their body. The film—as with all audiovisual images—materializes in the bodily sensations of a spectator:

> "Following Sobchack "[…] any movement occurring in the composition unfolds outside the physical film material, outside the screen. It is materialized as bodily sensation of/in the viewer, it is embodied by that viewer (Sobchack, 1992, p. 9) […] The Expressive Movement articulated in the medium of cinematic movement-images gains its affective reality as viewer's physical sensations. […] And it is exactly in this cinematic dimension of

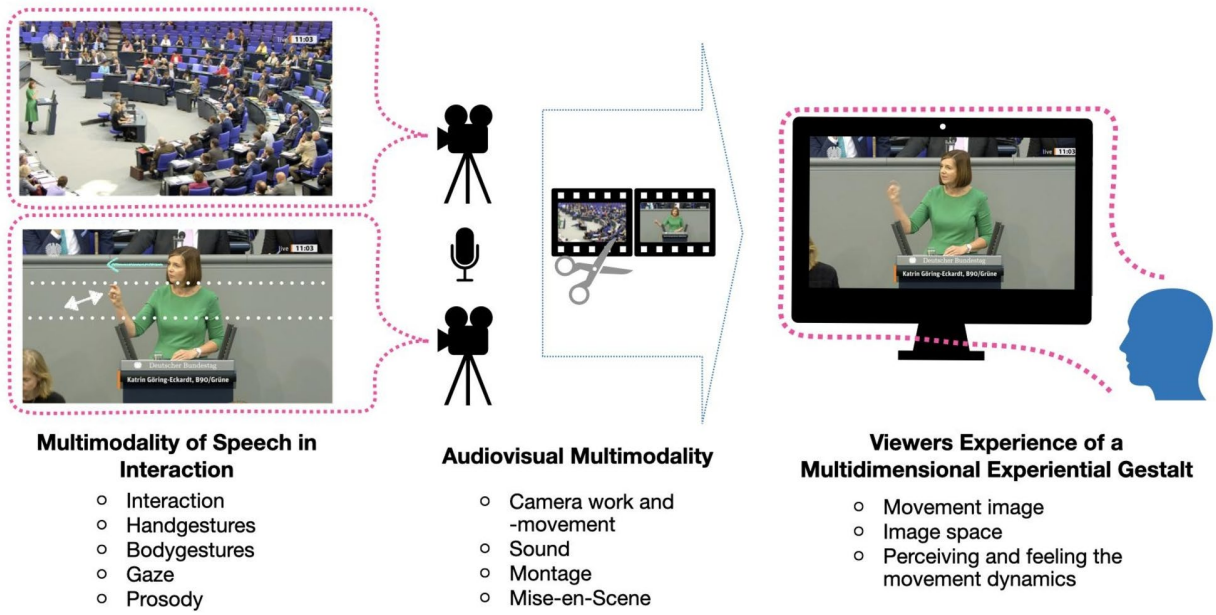


FIGURE 2
Two forms of multimodality forming the multidimensional experiential gestalt in the viewers' perception. Recordings from parliamentary debates are sourced from Deutscher Bundestag.

> movement, through which all constitution of meaning goes, where Sobchack finds the basis for the intersubjective dimension of cinema": (Müller and Kappelhoff, 2018, p. 64f).

This is how cinematic expressivity frames political debates affectively: viewers experience audiovisual images as bodily experience. If we connect this with Stern's idea of forms of vitality as characterizing face-to-face interaction as much as the cinema, music, or theater, then we get a feeling for the fundamental impact cinematic staging of affectivity may have on how multimodal speeches are felt and understood by spectators at their screens, namely as stance.

This is why we consider mediatized affective stance to be orchestrated by cinematic Expressive Movements. They modulate the audiovisual-represented verbo-gestural Expressive Movements of a given speaker, thereby mobilizing the affective experience of the viewers. As a consequence, we distinguish two aspects of multimodality in our analysis: the multimodality of speaking and audiovisual multimodality. We look at how they are intertwined, and how the orchestration of the audiovisual image (e.g., a close-up versus a wide-angle shot) frames the perception of the spectators of the political speech on screen.

3 Data and methods

3.1 The data

The full corpus of the project includes speeches from the 2019 and 2020 financial debates of the *Polish Sejm* and the *German Bundestag*. Audiovisual recordings were obtained from the official archives of both parliaments[1, 2, 3]. These recordings are freely accessible to any interested citizen.

3.1.1 The recordings

The Polish speeches were recorded on 8 January 2020 and the German ones on 10 and 11 September 2019. The German material consists of two debates (15 plus 23 speeches) with a total of 38 speeches of varying lengths (02:33–35:08 min). The Polish recordings contain 69 significantly shorter speeches (00:51–26:08 min). However, in total the length of the recorded material did not differ significantly: 323 min for the German and 339 min for the Polish videos. Each debate begins with the longest speech and ends with the shortest. In both parliaments, time limits are assigned beforehand. Speech slots are distributed based on political position and party affiliation. Each speaker is summoned by the President of the parliament and walks from their place to the rostrum. In both settings, the seat of the President and the seats for higher-ranking officials are positioned behind the speaker. The audience and other parties are arranged in a crescent tribune facing the speaker. In the *Bundestag*, camera shots show the speaker as well as the audience from the upper left, upper right, and upper front angles. In the *Sejm*, speakers are typically shown using a distant wide-angle camera when approaching the rostrum and a medium shot during the speech. In both parliaments, the footage is usually interleaved with the views of the audience, especially at the beginning and at the end of speeches and sometimes during applause.

3.1.2 Data storage

For data storage, exchange, and annotation management, we used the Corpus Mini client–server database system (Karpiński and Klessa, 2018, 2021, pp. 83–93). The Corpus Mini management system facilitates the monitoring of annotation and analysis workflow in multimodal corpora containing various kinds of multimedia files, annotations, and metadata materials. The present dataset includes. WAV and. MP4 files along with annotation files in the formats supported by Praat (Boersma and Weenink, 1992–2024), ELAN (Wittenburg et al., 2006), and Annotation Pro (Klessa et al., 2013). In Corpus Mini, each collection of files related to one recorded situation is treated as one data bundle named a recording session. Interoperability between annotation formats is ensured due to the implemented import–export options, which are crucial for the present context that requires the inspection of multimodal interactions based on measurements and annotation mining involving multiple parameters and annotation layers.

3.2 Methods

The methods outlined in this section bring together an empirical media-aesthetic (descriptive, phenomenological) approach with an instrumental analysis of speech prosody and a linguistic analysis of gestural performance.

They are an extension of the film-analytical and linguistic methods initially developed for the analysis of Cinematic Metaphor (CinMet). CinMet.

> "[…] addresses the temporality of meaning-making as a specific mode of perceiving, sensing, and feeling and offers different forms of visualizations of this temporal affectivity and the dynamics of […] meaning[−making]. Our starting point is the temporality of experiencing which characterizes film-viewing as much as face-to-face interaction." (Müller and Kappelhoff, 2018, p. 227).

Note that the term 'cinematic' refers here to the kinematic or 'movement' nature of both—face-to-face interaction as 'dance' in Stern's sense *and* film-viewing as an interactive process between 'movement-images' and spectators (Müller and Kappelhoff, 2018). Based on this earlier work on affectivity and multimodality, we approach stance-taking as it manifests itself in Expressive Movements. Again, we draw here on earlier study (eMAEX, electronically mediated analysis of Expressive Movements Kappelhoff et al., 2016; Müller and Kappelhoff, 2018, Appendix). For the analysis of moments of affective stance-taking in parliamentary speeches, we have focused on Expressive Movements that display a high affective engagement. These were described in terms of the movement qualities or vitality contours of the perceived movement gestalt and their multimodal orchestration.

1 https://www.bundestag.de/mediathek?videoid=7387900#url=L21lZGlhdGhla292ZXJsYXk/dmlkZW9pZD03Mzg3OTAw&mod=mediathek [Accessed 12 July 2024].

2 https://www.bundestag.de/mediathek?videoid=7387986#url=bWVkaWF0aGVrb3ZlcmxheT92aWRlb2lkPTczODc5ODY=&mod=mediathek [Accessed 12 July 2024].

3 https://www.sejm.gov.pl/sejm9.nsf/transmisja.xsp?documentId=D3D275F09D241460C12584EA004331A0&symbol=WYPOWIEDZ_TRANSMISJA.

To ensure intersubjective accuracy, analyses were carried out in independent tandems of German and Polish researchers. To complement the intersubjectively qualified analyses, instrumental measures of the basic prosodic properties were carried out. A moving time window approach was applied to capture changes in pitch frequency, intensity, speech rate (based on segmental duration), and normalized PVI. Time-group analysis was carried out for interpausal units identified in the realization of the EM.

3.3 Analytic procedure: affectivity as a form of stance-taking

Figure 3 shows an overview of the analytic procedure. The procedure falls into three larger steps: (1) setting the stage for analysis, (2) analysis of EMs, and (3) analysis of affectivity as stance.

3.3.1 Setting the stage for analysis

Analysis begins with a first preparation of the recordings (Figure 3, 1.1): It includes a transcription in the annotation software ELAN (Wittenburg et al., 2006) and Annotation Pro and the segmentation of the speeches into interactive units. Interactive units are bounded either by applause or by audible or visible interjections that receive a reaction from the speaker (the reaction can be minor such as raising the voice or a change in the flow of speech). In addition, making conversation inventories provides a first descriptive understanding of 'what's going on' in these interactive units.

Then, Areas with High Affective Engagement (AWHAE) are analyzed (Figure 3, 1.2), starting with annotating affective peaks and then distinguishing between areas with low and high affective engagement. The "identification" of the peaks does not imply an adding up of individual expressive modalities but rather follows the approach of temporally unfolding multidimensional expressive forms. They emerge as clearly perceptible changes in intensity from the so-called

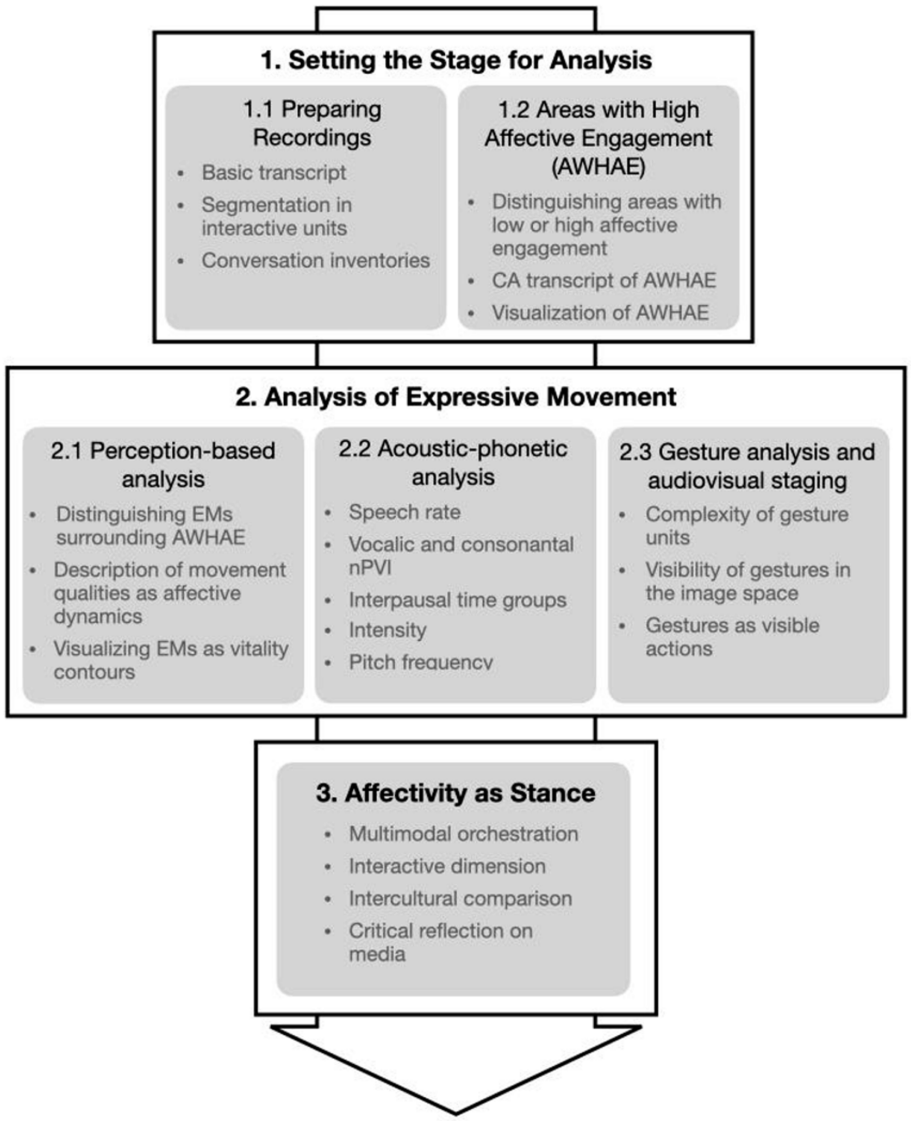


FIGURE 3
Analytic procedure: (1) setting the stage for analysis, (2) analysis of expressive movement, and (3) describing affectivity as stance.

baseline, i.e., the basic rhythm of the speech, as, e.g., explosive, steadily rising, or suddenly falling qualities. Baseline refers to the specific, individually characteristic speech style—or the multimodal dynamics of speech, from which the peak stands out. Affective peaks and baselines can only be understood as interrelated qualities of multimodally orchestrated speech. To prepare for micro-analysis of gestures and speech, a conversation analytic transcription of the AWHAE is made.

3.3.2 Analysis of EM

In the second step, the analysis of EMs begins. The concept of Expressive Movement serves as an analytical tool to capture an affective stance from the point of view of the perception of the viewers and listeners of the speeches on a screen. This means we consider the intertwining of the multimodality of speaking (as orchestration of speech and gesture) with the staging of it as audiovisual performance (audiovisual multimodality). The analysis of EMs in these speeches falls into three further analytic aspects: a perception-based analysis (2.1), an acoustic-phonetic analysis (2.2), and an analysis of gestural performance and audiovisual staging (2.3).

3.3.2.1 Perception-based analysis

In the perception-based descriptive analysis of EMs, the multimodal interplay of gestures, speech, and audience interventions in its dynamic affective unfolding is at stake. To capture the temporal unfolding of the affective dynamics, AWHAE are depicted as red triangles and blocks (using Inkscape as a tool). The EMs in which these affective peaks are embedded are shown as contours section 4.1).

Sometimes EMs may include two or more affective peaks, sometimes only one. Their different shapes are the different forms of perceived and experienced affectivity. In the next analytic step, these affective qualities of the EMs are described in terms of their movement qualities. We use Stern's list of adjectives as inspiration for the description (Greifenstein, 2020).

As in Stern's descriptions of vitality contours, the affective dynamics of EMs are variable. Expressivity changes as the speeches unfold and evolve in their interplay with the audience and depending on the camera angle and montage (the perceived audiovisual expressivity differs according to whether we see a speaker close up or in a wide-angle shot). It is this kind of affective expressivity that we try to capture and describe from various angles and in a combination of visualization and subjective verbal description. We seek to capture stance as affectivity as embodied in these multidimensional experiential gestalts *as the audience on the screens* perceives them. Agreement between researchers was extremely high—although we did not perform interrater reliability tests.

3.3.2.2 Prosody analysis

The next step addresses the dynamic contours of EMs with an instrumental analysis of the prosodic features. This analysis provides an independent source of 'evidence' for the perception analyses of the EMs.

Emotional prosody is placed very close to gestures in terms of expressive potential and conveyed meanings (Gibbon, 2005, 2011). Similarly to gestures, emotional prosody in adults may be stylized, modulated, or filtered by socio-cultural factors (Abelin and Allwood, 2000; Scherer, 2003; Scherer et al., 2011). In the study of emotional prosody, the acoustic-phonetic approach aims to identify and instrumentally measure acoustic correlates of prosodic features associated with emotionality. The results are often juxtaposed or coupled with perception-based experiments or expert listening judgments to test their relevance to human perception and the context of language communication. In the present study, several prosodic features have been identified in the expression of attitudes and emotions. We measured, in particular, changes in speech rate, normalized PVI, basic TGA parameters (slope and intercept), intensity, and pitch frequency.

Instrumental measurements of a selection of prosodic parameters pertaining to pitch, duration, and intensity were conducted using Praat (Boersma and Weenink, 1992–2024) and Annotation Pro (Klessa et al., 2013). Pitch frequency (in semitones, relative to the base frequency of 1 Hz) and intensity (in dB) were quantified utilizing respective Praat algorithms, wherein pitch extraction parameters were tailored to the speaker's gender. The resultant pitch and intensity tiers were then imported into Annotation Pro and utilized as input values for a moving average algorithm (cf. Karpiński et al., 2014), employing a 2-s time window with a 1-s time step. The moving average approach served as a means of smoothing the raw data to mitigate random fluctuations or abrupt changes in the measured values. To capture local variations in segment duration, normalized Pairwise Variability Index (nPVI) (Grabe and Low, 2002) was automatically computed for vowels and consonants separately, employing a plug-in for Annotation Pro (Klessa, 2016). Furthermore, we investigated syllable timing variability within pause-delimited stretches of speech (interpausal time groups), following the time group analysis (TGA) approach proposed by Gibbon (2013). We used the Annotation Pro implementation of TGA (Klessa and Gibbon, 2014). The TGA method uses the linear regression function of syllable durations, mainly for the regression slope values, as a first approximation to examining speech acceleration and deceleration patterns. The mean slope values calculated over large data sets were observed to serve as a potential indicator of speaking style for several languages (e.g., Gibbon et al., 2014). Discussing our results, we refer also to the baseline values of the above variables, i.e., the values measured in a selected stretch of speech (*ca.* 12 s) not belonging to any EM and recognized by experts as typical of a given speaker, emotionally neutral.

All the semi-automatic measurements were scrutinized for potential issues that might have resulted from the features of the recordings (e.g., noise and voices from other speakers). As the authors did not control the recording conditions, measurements may be, in general, less reliable, especially in the case of intensity where the distance from the microphone and speaker's orientation may bring in significant changes in the signal level. Furthermore, popular audio processing techniques such as noise gates and compression may distort the image of the original signal's dynamics. It is worth noting, though, that other prosodic features used in the study (duration and pitch frequency) are less sensitive to these aspects of recording quality, and that time-related analyses underwent double revision.

Obviously, there is more to how multimodal EMs are orchestrated than speech prosody and gesture, but in the moments of high affectivity that we are particularly interested in, these two appear to play an important role. In particular, when considering the audiovisual staging of the speeches, it appears that the complexity and visibility of gestural movements feature prominently in how affectivity as stance becomes a felt perception of political performance for online spectators of such events.

3.3.2.3 Gesture analysis

Gesture analysis takes into account how gestures are performed during EMs. Different aspects of gestural performance may become

relevant, and the micro-analysis of gestures is adapted accordingly (Müller, 2024). In the micro-analysis carried out for this paper, we consider the complexity, the visibility of gestural movements, and their character as visible actions in the Kendonian sense (Kendon, 2004) as they unfold over the course of an EM. In terms of their complexity, gestures are considered as temporal forms, which may be segmented into gestural units, phrases, and phases (Kendon, 2004, chapter 7) (Figure 5). Gesture units appear in simple and complex forms and are delimited by rest positions (Kendon, 2004, p. 111ff). Simple and complex gesture units differ regarding the number of gestural phrases that unfold along such a gestural unit of expression. Kendon's notion of gesture phrase includes preparation and stroke (cf. Kendon, 2004, p. 111ff.). Occasionally, a simple gesture unit may also be internally complex, namely when a gesture stroke unfolds with a complex movement pattern.

In addition to the unfolding complexity of gestural movements, we also considered the visibility of the gestural performance. Here, we considered in particular the height of their execution in gesture space. In gesture studies, McNeill's (1992, p. 89) systematics of different gesture spaces is an important analytical reference point. They concern, for example, the intersubjectivity of the gestural movements: a gesture that is placed high up in the gesture space moves into the central field of visual attention of co-participants. Such a gesture is foregrounded because it cannot be overlooked by attending partners in a conversation (Müller, 2008; Müller and Tag, 2010). To account for the visibility of gestures in moments of high affective engagement, we considered the gestural movement in its path and location through the gesture space.

As the execution of a gesture often involves the movement of the hand through several gestural zones, we took into account only the moment of the gesture where the hand is furthest from the starting point. The reference point for describing the height of gesture execution was the speaker's body. Gestures executed at the level of the abdomen were marked as 1, at the lower chest level 2, at the upper

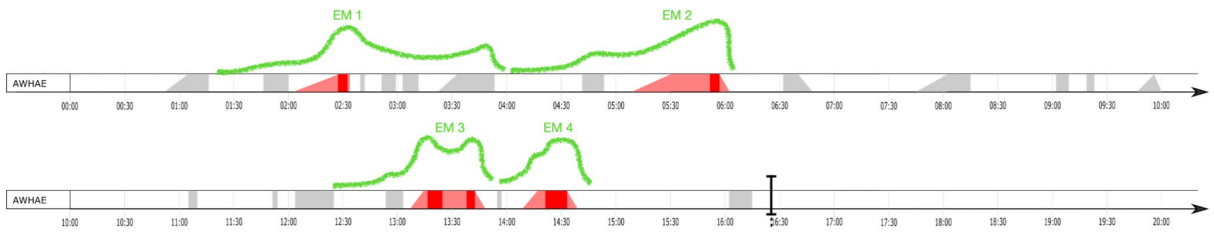


FIGURE 4
Bundestag speech: affective peaks along the speech (high affective engagement in red, lower in gray).

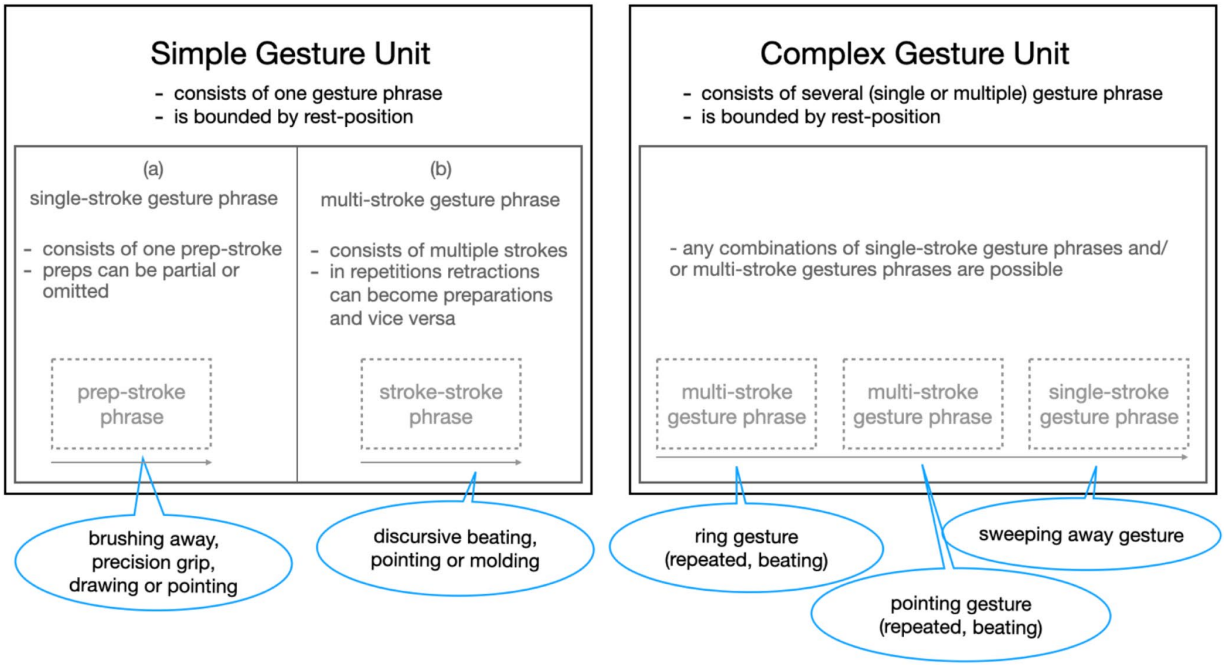


FIGURE 5
Simple and complex gesture units.

chest level 3, at the shoulder level 4, at the head level 5, and above the head 6.

In particular, the visibility of the gestures in audiovisual documentations depends crucially upon another factor: the orchestration of the audiovisual image space. Here is where audiovisual multimodality frames and shapes what of the gestural performances the spectators are able to see. That is why the EM analysis concerns the intertwining of the gestural performance with the orchestration of the audiovisual movement-image. We distinguish the following shot sizes: close-up, close shot, normal shot, medium shot, long shot, and wide angle (Figure 6, right).

Combining instrumental and descriptive analyses of EMs offers an alternative window on what otherwise could be easily dismissed as 'only' subjectively perceived movement gestalts. Because prosody is a movement that makes a sound (Gibbon, 2005, 2011), the analysis of prosody is an interesting complement to the perception-based description of multimodal EMs, as both speech prosody and gesturing are body movements that contribute to a multimodal, multidimensional gestalt that we conceive here as one EM.

Describing affectivity in terms of stance, box 3 of Figure 3, is discussed in detail and with reference to the analysis of expressive movements in the following and in section 5.

4 Analyzing affectivity as stance in parliamentary debates

For this article, a close analysis of one speech from each of the German and the Polish debates serves as case studies to exemplify the media-aesthetic approach to multimodal affective stance-taking. This analysis exemplifies how multimodal EMs may serve as a frame for the investigation of affective multimodal stance-taking in recordings of parliamentary speeches. In the following, a detailed analysis of one expressive movement with an area of high affective engagement from the Polish and the German material is offered. We begin with a perception-based, descriptive account of the affective qualities of the EM, complement this with an acoustic-phonetic analysis of the parts of the speech that entail this EM, and end with a micro-analysis of the gestural performances and the audiovisual staging of the two EMs.

4.1 Bundestag: the speech and speaker

The speech is given by Katrin Göring-Eckardt (KAGE), the parliamentary leader of the *Green Party* (*Bündnis 90/Die Grünen*). Her speech has a total length of 16:13 min. It is the 6th speech of the second day of the financial debate with a total of 23 speakers and a length of 3:45 h altogether. The budgetary and the financial debates are part of the regular agenda of the chancellor. Chairpersons of the opposition challenge the budget of the government, whereas the chancellor and other members of the government hold speeches to defend it.

4.2 Sejm: the speech and speaker

This speech is delivered by Dobromir Sośnierz (DOSO), a representative from the right-wing conservative party, *Liberty and Independence Confederation* (*Konfederacja Wolność i Niepodległość*). Dobromir Sośnierz's speech lasted a total of 1 min and 8 s. It was the 70th speech in the entire budgetary debate, which featured a total of 135 speeches. The debate took place in a single day.

Both parties represent distinct views, most possibly having both passionate critics and supporters within the parliamentary chambers. The difference in speech length was not a factor here as the focus is on the micro-analysis of one EM.

4.3 Perception-based analysis of EMs: movement dynamics as affective qualities

When considering the speeches in their unfolding entirety, we see that they have an overall characteristic flow and dynamic. The micro-analysis focuses on moments of high affective engagement that stand out in particular. The multimodal EMs that surround them serve as units of analysis for the investigation of verbal and audiovisual multimodality.

4.3.1 First impressions of the Bundestag speech

The German speaker, Katrin Göring Eckardt (KAGE), is a very enthusiastic and engaged speaker, and she speaks with a constant high

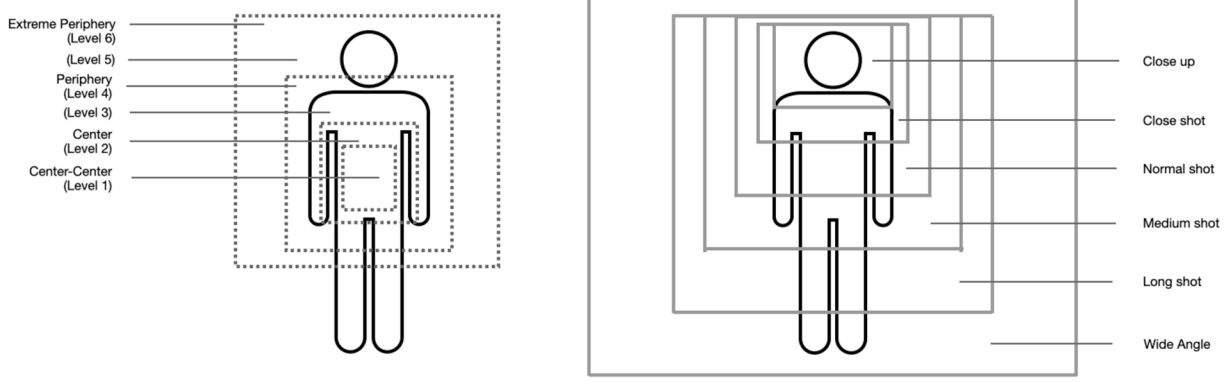


FIGURE 6
Gesture space and image space.

volume and pitch. Her speech is interrupted by various (harsh) heckling from the audience which she mostly ignores. At one point she is interrupted by the president of the *Bundestag* (*Bundestagspräsident*) Wolfang Schäuble because of a speaking request by another politician which she allows. KAGE uses a lot of discursive beating gestures. Her overall speaking style is very clear, sometimes—by pronouncing syllable after syllable—it is highly emphasized. Figure 4 shows the unfolding of affective peaks along the speech.

Overall the speech is characterized by a rather moderate affective dynamics. Most of the 23 peaks are moderate (in gray), and only five show high affective dynamics (in red). The movement dynamics of the affective peaks in general are rather calm, concentrated, and determined. The green vitality contours indicate the movement dynamics of the EMs that surround the affective peaks.

4.3.2 Affective dynamics: an expressive movement in the Bundestag speech

The affective dynamics described in this section concerns the last EM that shows high affective engagement in KAGE speech (EM 4, min. 14). It includes the audience's reactions, as we take the perspective of the viewers that follow the speeches on their media devices. The description follows the eMAEX system (Kappelhoff et al., 2016). We use musical terms and the vocabulary for vitality forms as suggested by Stern (2010).

> *In terms of its overall movement dynamics, the expressive movement is characterized by a steady crescendo that evolves into a peak phase with an enduring insisting character that terminates with a decrescendo including applause from the audience. This intense dynamics of the expressive movement begins rather quietly but then increases in a steady and pushing manner with increasing tension in voice and body. Speaking quickly and with a clear pronunciation the crescendo phase unfolds. During the peak phase, KAGE speaks with strong intonation and with a loud and tensed voice. While her gaze addresses the audience seated in front of her (this is where the Green Party members and the Christian Democrats are seated), she points repeatedly to the right in an accusing manner at the representatives from the right-wing party (Alternative für Deutschland, AfD). The rhythm of the peak phase includes significant pauses and fades away with a closing sentence that is drowned out by applause.*

4.3.3 First impressions of the Sejm speech

Dobromir Sośnierz (DOSO) is an extremely expressive speaker. His speech is characterized by immense emotionality, which is manifested both in his way of speaking and in his gestures. In his statement, DOSO presents his vision of the budget. The vocabulary he uses is very direct, sometimes even aggressive. DOSO directly addresses left-wing MPs, politicians from Law and Justice, and one specific MP, accusing them of lacking understanding of how the budget works. His bodily behavior also reflects his high affective involvement. We can observe full-body enactment, including leaning toward the audience in front of him and to his left and right-hand sides. DOSO performs numerous gestures, predominantly discursive beats. Some of them are performed with arms extended at the elbow. This affective dynamics shows in the visualization of affective peaks: in particular, almost his entire speech is one area with high affective engagement, embedded within one EM (Figure 7). Although the fact that DOSO's speech is rather short certainly plays a role, it is still remarkable that he maintains a high affective engagement all the way through his speech.

The area with high affective engagement in DOSO's utterance is highly dynamic, featuring heightened volume and pitch, sweeping upper body movements, expansive gestures, rhythmic beats, and animated facial expressions. During DOSO's speech, individual voices of parliamentarians can be heard. It seems that these voices neither influence DOSO's speech nor constitute comments on his statements.

4.3.4 Affective dynamics: an expressive movement in the Sejm speech

The EM in DOSO's speech is extremely dynamic and highly affective. Here, the area with high affective engagement is co-extensive with the EM:

> *The Expressive Movement starts rather calmly but quickly develops toward a rapid increase in intensity, marking the start of the peak phase: after continuously accelerating speech pace, gestural beats increase in size and space. Successive, staccato gestural beats follow along with an explosive peak with even bigger and more energetic, repetitive beats and a change in pitch quality. This is followed by a slight decrease in intensity.*

4.3.5 Summary and discussion

The perception-based descriptive analysis of the affective dynamics reveals that in both speeches affectivity unfolds in rhythmic patterns, sometimes variably and with changing affective intensity; in short speeches, high affective dynamics may even shape the entire speech. Visualizations and descriptions of the affective qualities reveal perceptive gestalts which are considered multimodal Expressive Movements. These EMs unfold as characteristic forms of vitality. Perception refers here to the perception of the speech as an audiovisual recording. Researchers, like viewers, can only perceive what the shot size and angle, the montage, and the recorded sound allow them to hear and see. Particularly relevant here is the role of the sound settings of the microphones, and whether they are set to strengthen the speaker's voice and background or suppress audience reactions or not. The multimodal EMs described here thus emerge as multidimensional experiential gestalts in the process of watching the speeches on TV or on some kind of screen.

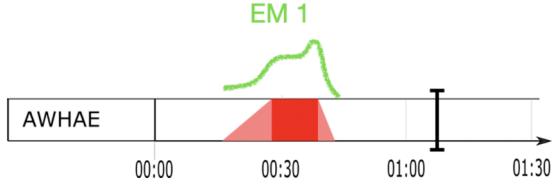


FIGURE 7
Sejm speech: affective peaks along the speech (only high affective engagement in red).

4.4 Acoustic-phonetic analysis of EMs

The instrumental analysis of the prosodic characteristics of the speeches was carried out as an exploratory study to investigate whether they would reveal similar forms of affective dynamics and thus support the descriptive analysis of EMs and their specific forms of affective dynamics. The subject of analysis is the sound of the recording, not the 'real' local speech sound happening *in situ*. This means also the acoustic-phonetic analysis also takes the perspective of the viewer and listener of the recorded speech.

The acoustic-phonetic analysis included the following variables: speech rate, vocalic and consonantal nPVI, syllable duration difference slope and intercept for interpausal time groups, intensity, and pitch frequency.

4.4.1 The speech from the Bundestag

As KAGE's speech is rather long and contains several moments showing higher affective dynamics, a moving time window approach was selected that zoomed in on the region within the speech that entailed the EM (4) described above.

The affective dynamics on the level of the acoustic-phonetic features of the speech, within the range of the EM, is displayed in Figure 8. It visualizes the variability of selected timing and pitch frequency parameters in the utterances of KAGE. The plots in Figure 8 represent the mean values for moving time windows within the EM range and refer to the *changes in the speech rate* (phones per second), *normalized Pairwise* Var*iability Indices* (*nPVIs*), *intensity* (dB), and *pitch frequency* (semitones re 1 Hz) (last 3 s are not shown in the graphs as they contain no speech). The points are joined with a cubic spline curve. Additionally, the fifth-degree polynomial regression lines (red) are added to the first three figures to highlight more general, long-term tendencies. Figure 8C illustrates the values of *syllable duration* difference regression slope (the blue plot) and intercept (the red plot) for 15 *interpausal time groups* within the realization of the EM. In all figures, the green graphs show the affective dynamics as vitality contour from the descriptive analysis of the EM above. The green line was manually superimposed on the software-generated graphics.

In particular, the dynamics of the prosodic features within the EM differ from what is observed within the baseline dynamics of the speech. The differences do not occur only between central tendency or variability measures but also between the (complex) patterns of their changes over time. Moreover, as shown below, they are of different nature for different prosodic features, and they do not necessarily occur in all of them simultaneously. The speech rate within the EM [mean rate = 10.87 phones/s, SD (standard deviation) = 3.89] is lower than the baseline (mean rate = 11.99 phones/s, SD = 2.69), but the higher standard deviation value may indicate more dynamics in the changes of speech rate within the EM. The value of the vocalic nPVI within the EM is lower than the baseline (mean nPVI = 61.57, SD = 14.16 vs. mean nPVI = 55.08, SD = 14.16) which may mean a higher rhythmic stability, perhaps due to the strong, emphatic rhythm of speech. In the TGA, the mean of slope values for the EM (mean slope = 0.0362) is higher than the baseline (mean slope = −0.0099), indicating the tendency toward deceleration within the EM. The mean intercept value for EM (mean intercept = 202.68) was higher than the baseline (mean intercept = 198.63), which indicates a slightly slower average tempo and longer syllable durations in the EM. Speech rate and vocalic nPVI curves are complex, and their correlation with the vitality curve is not obvious. However, one may notice that the affective peak is preceded and ended with a speech rate of especially high ranges while the vocalic nPVI seems not only to gain more fluctuations after a short period of stability but the fluctuations show more periodic regularity.

In addition, the intensity is higher within the EM (mean intensity = 66.10 dB, SD = 5.41) than the baseline value (mean intensity = 65.20 dB, SD = 6.26). In this case, the lower standard deviation, corresponding to "flattened" dynamics, may be a result of an attempt to keep the voice louder all the time. As mentioned, however, it may be also due to compression used in voice amplification. The pitch frequency in the baseline is *ca*. 0.5 ST higher than the mean for the entire EM (97.78 ST with SD = 2.81 vs. 97.22 ST with SD = 3.08), but, as in the case of speech rate, a higher value of standard deviation indicates a higher spread of the values, i.e., operating on a wider range of pitch levels.

Both the intensity and the pitch frequency changes, in spite of high variability, follow the general shape of the vitality curve, increasing toward the affective peak and remaining at a high-level plateau almost till its end. As shown in the measures discussed above and in respective figures, the prosodic means of expression used by KAGE are complex and vary in the unfolding of the EM. The most remarkable feature is the increase in pitch frequency, highly correlated with intensity changes (Figures 8D,E), including the compression of the intensity curve in the affective peak region. Timing-related values, on the contrary, may suggest more prominent and regular fluctuations in this area, with a slight tendency toward deceleration.

Due to the relatively long duration of the EM and its complex internal structure, global measures of prosodic parameters may not fully reflect its dynamic nature. They may, however, be more evident for the affective peak when contrasted with the baseline characterized by the overall dynamics of the speech. The mean values of speech rate, intensity, and pitch frequency within the affective peak turned out to be higher than in their mean values within the entire EM as well as the baseline values (mean speech rate = 12.71 phones/s, SD = 2.58; mean intensity = 66.97 dB, SD = 4.67; mean pitch 99.51 ST, SD = 3.19) while the mean vocalic nPVI was still lower than the baseline value (mean nPVI = 60.69, SD = 18.01), and the mean slope is higher than the baseline but still lower than the mean value for the affective peak (slope = 0.0096, intercept = 204.08).

To sum up, the affective peak of the EM under analysis is easily distinguishable and characterized by distinct prosody. While the dynamic changes of prosodic features, initially predominantly increasing (but not necessarily in a stable, steady fashion), and decreasing in the final stage (again, this process may be quite complex as well), are typical of the EM as a whole, in KAGE, pitch frequency and intensity reach the highest values with a more limited degree of variability ("flatter" sections of the curves in Figures 8D,E), becoming prosodic indicators of the affective peak.

4.4.2 The speech from the Sejm

The speech from the Sejm is short and almost co-extensive with one EM that unfolds with a single affective peak (cf. Figure 7, section 4.1). Consequently, the instrumental analysis of the prosodic parameters extends over the entire speech. As in the German example, instrumental measures of selected prosodic parameters related to pitch and duration were taken and their values were confronted with

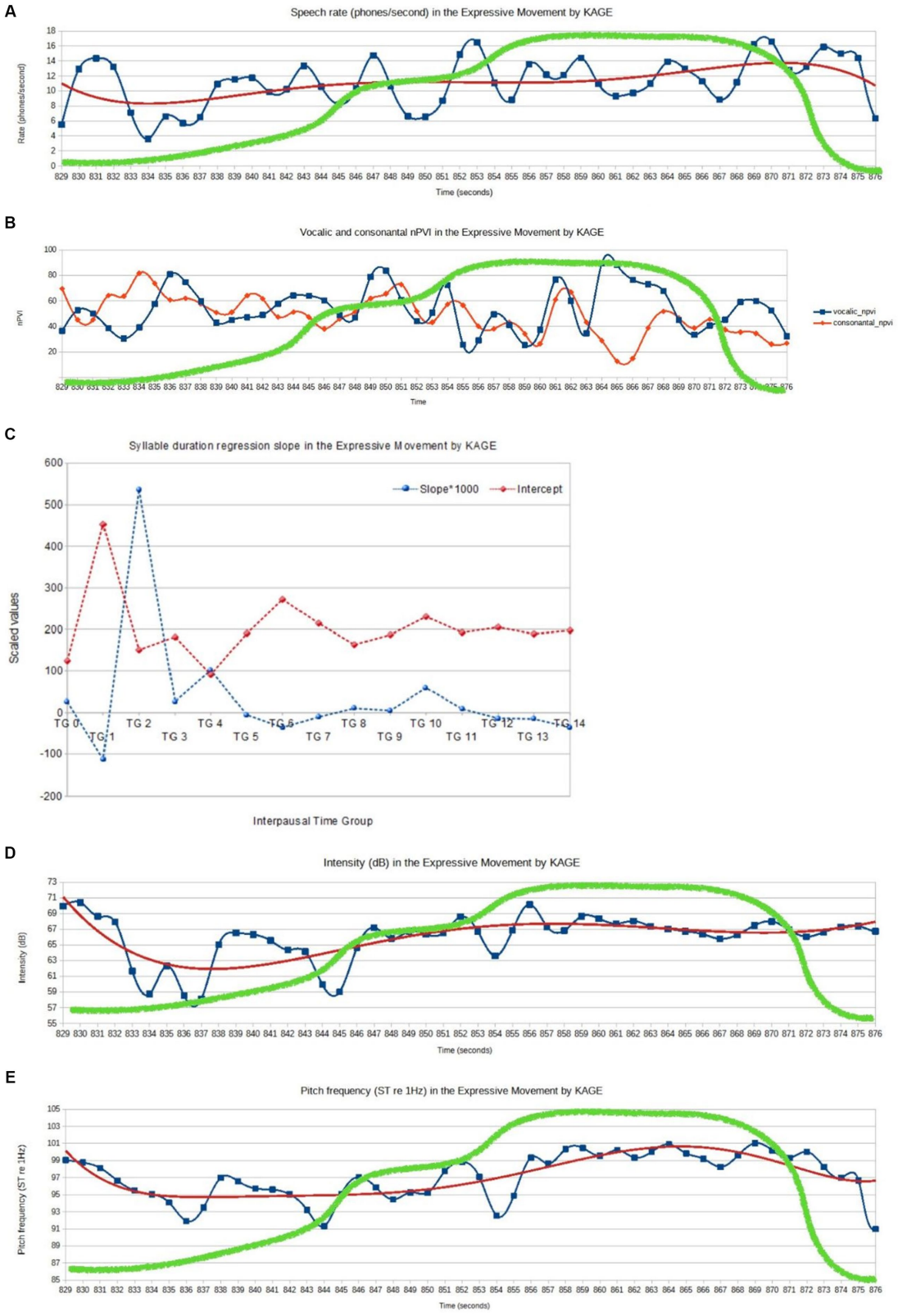


FIGURE 8
Acoustic-phonetic features of the speech and superimposed vitality graph (geen) from the descriptive analysis, within the range of the EM (KAGE). **(A)** KAGE, EM: speech rate (phone/second). **(B)** KAGE, EM: vocalic and consonantal nPVI (window size = 2 s, time step = 1 s). **(C)** KAGE, EM: mean values of syllable duration difference slope and intercept for interpausal time groups (TG stands for time group). **(D)** KAGE, EM: intensity (dB, window size = 2 s, time step = 1 s). **(E)** KAGE, EM: pitch frequency (semitones re 1 Hz).

the results of the perception-based analyses of the EM with its affective peak area.

DOSO achieves extremely high levels of prosodic prominence in the affective peak, but the entire realization of EM is characterized by dynamic changes in prosody. In Figure 9, the mean values calculated from a moving time window within the EM are represented for speech rate (phones per second), vocalic and consonantal normalized Pairwise Variability Indices (nPVIs), intensity (dB), and pitch frequency (in semitones re 1 Hz). A fifth-level regression line is added (red) to indicate long-term tendencies. Figure 9C represents the results of the Time Group Analysis, i.e., syllable duration difference regression slope (the blue plot) and intercept (the red plot). While the speech rate varies throughout the realization of EM, with a decreasing tendency, its mean value remains very close to the baseline (mean rate 13.92 phones/s, SD = 3.80 vs. mean rate = 13.99, SD = 3.82). The mean vocalic nPVI is higher in the EM than the baseline, with an even more noticeable difference between the standard deviation values (voc.nPVI = 58.06, SD = 17.79 vs. voc.nPVI = 53.65, SD = 12.00), indicating a higher spread of the values. Time Group Analysis indicates that more duration regression slope variability occurs closer to the end of the EM, exhibiting more tempo differentiation in the final part of the EM. The mean intercept value for the EM is higher than for the baseline (intercept = 139.08 vs. intercept = 135.54). The values of the duration-related variables suggest that in the initial phase of the EM realization, the speaker has a decreasing tendency while closer to the affective peak, there are strong rhythm fluctuations. The periodic changes of the nPVI values become remarkable immediately before the first rise of the vitality curve, but they do not stop immediately after the affective peak. The mean intensity in the EM is very close to the baseline (66 dB, SD = 1.7 dB vs. 65.7 dB, SD = 2.04 dB). As in the case of the German speaker, it may be partially caused by compression but DOSO seems to speak at a high-intensity level from the very beginning of his speech. On the other hand, the difference between both the mean pitch value and its standard deviation in the EM and the baseline is striking (89.85ST re 1 Hz, SD = 5.45ST vs. 83.73ST, SD = 1.06ST). The mean pitch almost continuously rises until it reaches a plateau within the area of affective peak, with another peak after the vitality curve is steeply falling down. The relationship between pitch frequency and intensity is not as obvious here as in the German speaker. Intensity seems to initiate rapid prosodic changes just before the occurrence of the affective peak, with the pitch value changes following shortly afterward. The fluctuations typical of the intensity values within the affective peak got weaker gradually after the vitality line abruptly headed downward, while the pitch values gradually fell down as well with a certain delay.

The pitch change seems to be the dominant means of expressivity used by DOSO in his speech. It rapidly rises and remains higher throughout the EM to fall only at its end. Duration-related measures do not bring evident differences. They may require a quantitative exploration of a larger dataset to find more subtle tendencies.

The affective peak shows even more stable prosodic characteristics than in the German speaker. Pitch height reaches a plateau at a very high level (mean pitch = 95.13ST, SD = 1.55), and speech rhythm becomes regular and prominent, but the nPVI values (voc. nPVI = 63.44, SD = 14.64) grow probably due to the forced adjustment of the segmental durations to the strong rhythmic pattern. Similarly, the results of the TGA indicate significant changes in the second part of the EM. However, the speech rate is preserved, and not very different from the baseline value (speech rate = 13.56, SD = 3.05).

Similarly to what was observed in the German speaker, the area of the EM is characterized by more prosodic variability than the baseline. However, it is worth mentioning that the changes do not simultaneously occur in all the variables. Some of them, like the speech rate in DOSO, may remain almost unchanged or change in a way that cannot be properly captured by global measures like the mean or the standard deviation. In DOSO, extremely high pitch frequency remains the most prominent component of the affective peak prosody.

4.4.3 Summary and discussion of the two speakers

The German and Polish speakers used for the demonstration of the present framework of analysis differ in gender, their political views, their way of speaking, and the duration of their EMs under analysis vividly differ as well (*ca*. 50 s for the German and *ca*. 26 s for the Polish speaker). Although both of them are fluent and expressive speakers, their "prosodic profiles" are very different. The speech rate within the EM in DOSO gradually decreases, while in KAGE the overall tendency is increasing. What they share, however, is strong fluctuations that seem to reflect perceivable peculiar rhythmic patterns. However, speech rate seems to be a secondary resource for emotional expression for the recordings under analysis. In both speakers, the initial part of the EM is characterized by lower fluctuations of nPVI (both vocalic and consonantal). This may suggest that the rhythm of speech becomes more prominent later in the EM, especially in the affective peak. Speech acceleration and deceleration phenomena were observed in both speakers. However, again, they manifested different strategies with regard to the variability of duration difference regression slope that corresponds to the average rate of duration change in the data. While in the realizations of the EM by KAGE more differentiation is visible at the beginning of the EM, for DOSO, larger differences between the subsequent time groups were found closer to the end of the EM. In the case of DOSO, the slope changes appear to quite strongly coincide with other prosodic features and subjective judgments of the speaker's affective involvement, whereas for KAGE, this relationship is not obvious. Intensity also changes in different ways in the two speakers. In DOSO, it remains on a relatively low level, and only at the beginning of the affective peak, it abruptly grows, and later, within the affective peak, it strongly fluctuates, while in KAGE more fluctuations can be seen at the beginning of the EM, and later, she seems to operate on a higher level, with a more stable intensity values. Finally, the usage of pitch is clearly different in the initial part of the EM, where DOSO speaks with a relatively low voice and with limited pitch fluctuations, while the fluctuations in KAGE are much deeper than within the affective peak area. DOSO's pitch frequency reaches a very high level in the affective peak, forming a "plateau" and falling down after the end of the affective peak. The pitch frequency values in KAGE's affective peak also seem to be "compressed" and more stable, in this case, than at the beginning of the EM.

In many respects, different in the German and Polish speakers, especially referring to the temporal development of the EM, what is particularly important is that the dynamics of the prosodic features under scrutiny are? The above-mentioned changes in pitch in the section before the affective peak may be one example: in KAGE, pitch

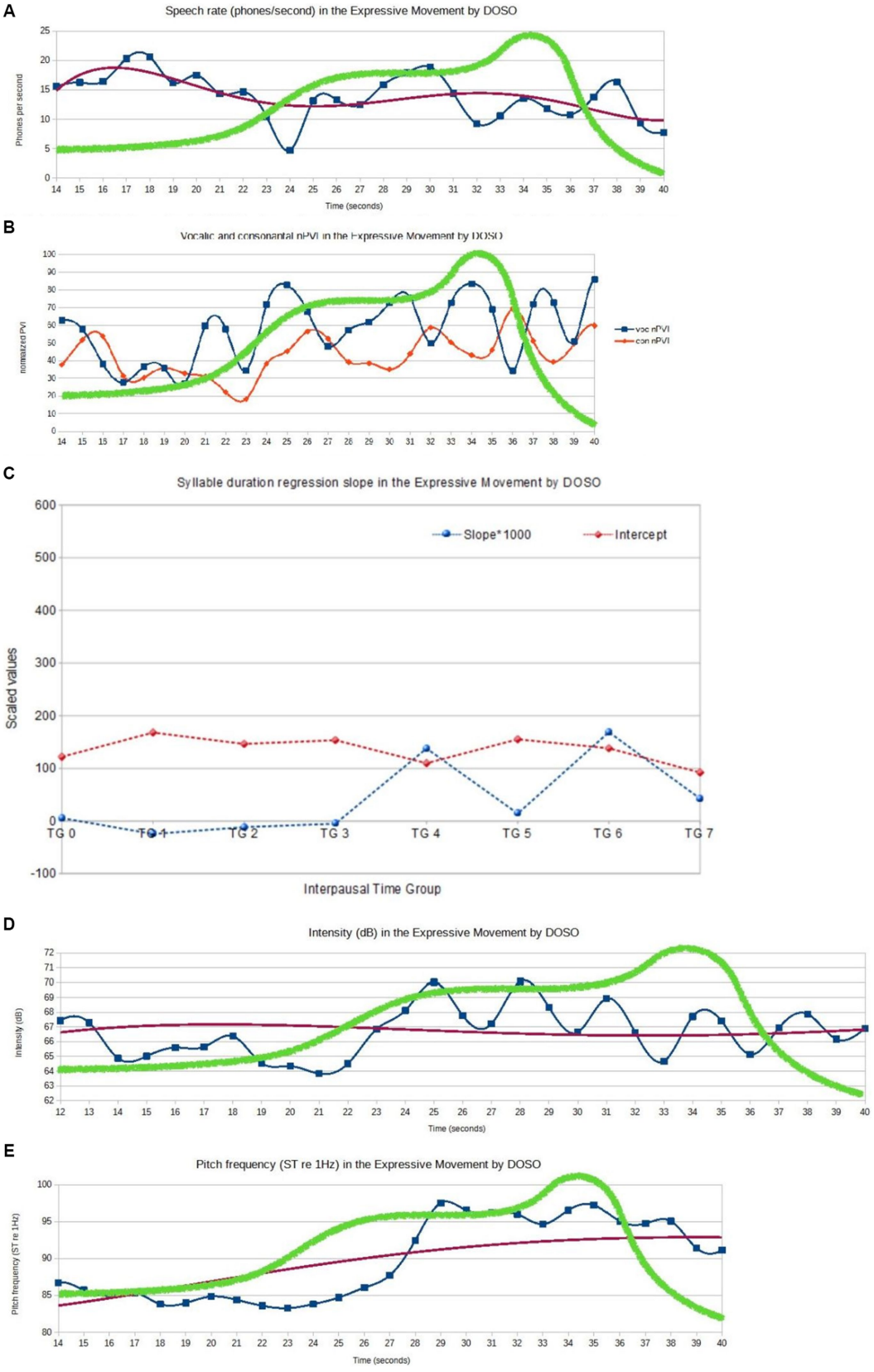


FIGURE 9
Acoustic-phonetic features of the speech and manually superimposed vitality graph (green), within the range of the EM (DOSO). **(A)** DOSO, EM: speech rate (SRMA, phones/second). **(B)** DOSO, EM: vocalic and consonantal nPVI. **(C)** DOSO, EM: mean values of syllable duration difference slope and intercept for interpausal time groups (TG stands for time group). **(D)** DOSO, EM: intensity (dB). **(E)** DOSO, EM: pitch frequency (in ST).

frequency changes are "dynamic," while in DOSO, the pitch trace is flatter, with relatively low pitch frequency values.

While the dimensions of prosody are not fully independent, it should be noticed that the prosodic features potentially typically related to the expression of emotionality seem to be used selectively by both speakers. In KAGE, for example, pitch frequency and intensity highly correlate, but speech rate changes have clearly different dynamics. In DOSO, the similarity of pitch frequency and intensity traces is less obvious, and duration-related variables develop in a different way. Combining certain subsets of such features may be partially due to the physiology of speaking, to the properties of a given language, but can also be a part of the more or less conscious expressive strategy. Full analysis of their contributions and interactions will be subject to future work as it goes far beyond the scope of the present study.

What the results indicate, however, is that there is a marked difference in the acoustic-phonetic level between a baseline of the speech and the EM, which is particularly apparent in the affective peaks. The domains of pitch and intensity seem to be used for expressing emotionality more extensively by both the speakers. Pitch frequency and intensity reach higher values and seem to be compressed, with plateaus in the region of the affective peak. Duration-related measures provide evidence of a different kind, showing stronger periodic fluctuation. The instrumental analysis thus supports the perception-based recognition of EMs as perceptive gestalts.

4.5 Gesture analysis and audiovisual staging: complexity, visibility, and gestures as visible actions

This section considers the gesture performance as another significant facet of the multimodal orchestration of EMs with high affective dynamics in the parliamentary speeches. It begins with a general description of the multimodal orchestration and then addresses pertinent aspects of the respective performances more systematically and in more detail: in particular, complexity, visibility, and gestures as visible actions.

4.5.1 Bundestag: multimodal orchestration of the EM

The EM unfolds around interjections from members of the right-wing party AfD who react to the warning of KAGE that the collaboration of AfD and the Christian Democratic Party CDU would be highly dangerous for democracy. KAGE then describes the members of the right-wing party metaphorically as 'arsonists' who subvert and destroy democracy and therefore need to be stopped by the 'democrats'. The interjections from the right-wing MPs in the audience work as a boost for the affective dynamics of the speech. KAGE begins her speech in a calm and concentrated voice and subtle hand movements transforming increasingly in higher volume, speech velocity, and large gestures. Pragmatically, the EM begins with a warning but then changes to an accusation and a defense culminating in a clear positioning of KAGE as part of the 'democrats'. This change in pragmatic meaning-making concerns both the verbal and the gestural part of the multimodal utterance.

Whereas the hand form in the beginning is rather non-specific and the movement is a subtle trembling, they change to a complex sequence of highly articulate specific gestural forms, with a shift to increased volume and speech rate. Concurrently KAGE makes a precision grip and an index finger pointing. For everyone to see, they are positioned in the center of the gesture space and are moved up and down in accentuated and rhythmical beats.

In the moment of highest affective intensity, the speaker shifts her upper body and gestures toward the right side of the audience (where the right-wing party has their seats). With this change in body orientation, the gesture space moves to the right too, the MPs from the AfD concurrently become the addressees of repeated index finger pointings and a decisive/determined negating sweeping away gesture. With the very last phrase of the last gesture unit, KAGE embodies a gestural full stop: with an emphatic downward beat, performed with a both-handed pointing gesture she marks the end of the argument and the strong stance she is taking against the interferers.

In a nutshell, KAGE's change of affective dynamics within the EM is not only traceable on the level of the speech performance but also in regard to the communicative actions and the movement qualities of the hand gestures and body posture.

4.5.1.1 Gestural complexity

Within the approximately 1-min-long EM, KAGE produces a series of four gesture units (Kendon, 2004, chapter 7) that increase in complexity and visibility over time. The increase in complexity concerns the internal structures of the gestural units. While the first two gesture units in the EM are simple units with only one gesture phrase (Kendon, 2004, chapter 7), the following two gesture units consist of 13 and 3 gesture phrases (Figure 10).

Simple Unit (1) is formed by one multi-stroke phrase (Figure 10): KAGE makes small rhythmical beats on the rostrum that go along with speech rhythm and accentuation. Simple Unit (2) evolves as one multi-stroke phrase, here KAGE's gesturing changes to subtle, vibrating up–down movements remaining on the rostrum, anchored at the wrist of her right hand. With the third gesture unit, KAGE reaches the affective peak phase, which unfolds as a highly complex gesture unit (3). Alternating between multi-stroke and single-stroke phrases, a sequence of 13 gesture phrases unfolds. KAGE makes a series of rhythmical tappings on the rostrum and then changes to accentuated pointing beatings toward the right side of the parliament. Over this sequence she moves her gestures higher up in the gesture space, from the lower right to the center and eventually to the right periphery, concurrently using her arm to perform the accusing pointing movements (Figures 11, 12).

4.5.1.2 Visibility of gestures and audiovisual orchestration

The increase in gestural complexity goes along with the higher visibility of the gestures. During the affective peak, KAGE not only moves her hands up to the center of the gesture space, but they concurrently also become larger. Figure 11 visualizes the vertical coordinate of the stroke positions, showing a similar rising and falling pattern as the complexity of gesture units. This gestural performance makes the hands more visible not only for the attending audience in the parliament but also for the viewers on the screens, as the mise-en-scène of the audiovisual image places KAGE in the center of the image space and a medium shot perspective for the most part of the EM.

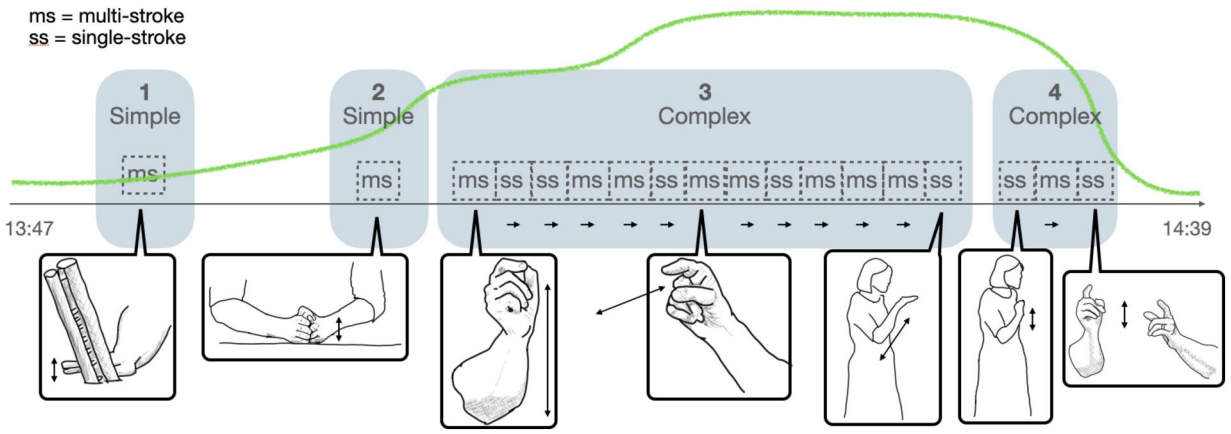


FIGURE 10
KAGE, the complexity of gestural performance.

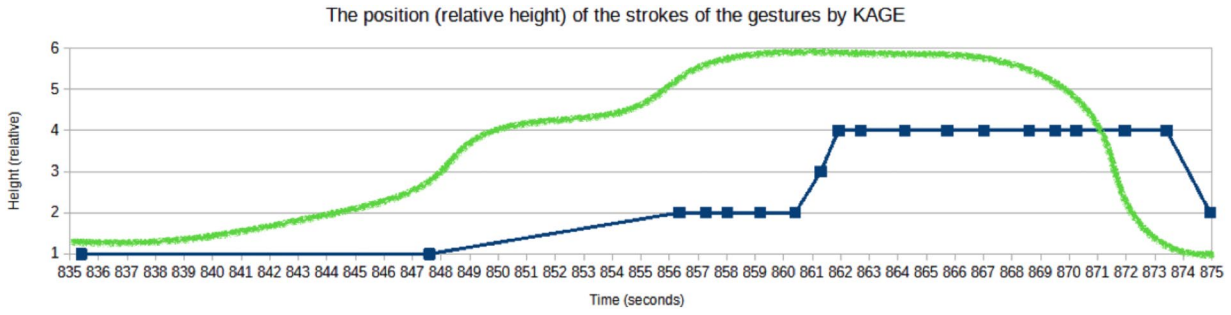


FIGURE 11
The vertical coordinate of the position of stroke realization by KAGE within the range of the EM (where "1" is the height of the abdomen and "6" is above the head) (Peak: from 861st to 874th sec).

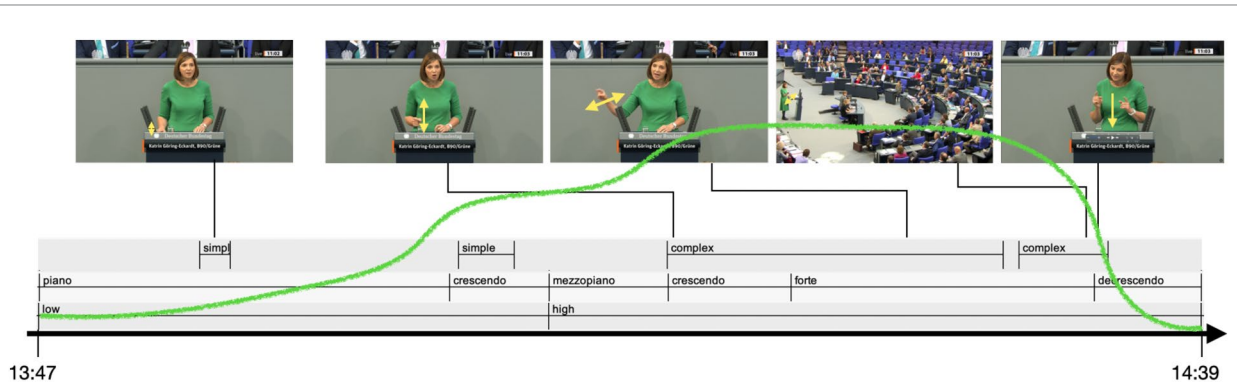


FIGURE 12
KAGE, vitality contour, musical terms, and the visibility of gestures in the image space as the EM unfolds. Recordings from parliamentary debates are sourced from Deutscher Bundestag.

The gesture units unfold an affective dynamics that can be described as a flow from crescendo (units 1 and 2) to crescendo and forte (unit 3) to decrescendo (unit 4) (Figure 12). Gestural complexity *goes along with* the high affective dynamics of the expressive movement. However, while her gestures are clearly becoming more visible to the audience in the parliament, this is not the case for the viewers of the audiovisual recording. Because a cut to a wide shot reduces the visibility of the gestures dramatically making a determined sweeping away gesture along with its expressive dynamics nearly invisible (Bressem and Müller, 2014) (Figure 12). The insisting prosodic qualities (pitch, intensity), however, remain perceivable for them. Then, with a camera-shift back to a medium shot there is an increase in visibility for the spectators at home, concurrently showing a series of single-, multi-, single-stroke phrase, e.g., a decrease in

internal complexity. This last gesture unit then contributes significantly to the termination of this EM: KAGE makes a kind of gestural exclamation mark, a highly accentuated, energetic both-handed downward pointing gesture (Figures 11, 12).

Thus, over the course of the unfolding of the EM, the gestural movements not only become more complex but also more visible: they become bigger, are moved higher up in the gesture space, and are shifted to the center of the image space (Figure 12). The visibility for the spectators changes when the camera perspective changes to a wide shot. This zooming away from the speaker's gestures decreases the perceivable expressivity of the gestural performance for the spectators of the televised performance. However, the audiovisual perspective shifts concurrently include the MPs and the accused audience, thus creating a visual image that foregrounds the interaction between the speaker and parliament.

Summing up, a closer look at the sequence of four gesture units shows a steadily rising affective dynamics materializing as increasing and decreasing complexity and visibility of gestural movements. Dynamic changes in gestural complexity participate in the multimodal orchestration of this area with high affective engagement. They are one facet of the multimodal orchestration of affectivity as stance. For the audience that follows such televised speeches, the audiovisual orchestration contributes significantly to the multimodal orchestration of an affective stance-taking.

4.5.2 Sejm: multimodal orchestration of the EM

The EM begins with a rhetorical question addressed to all the members of parliament, in which DOSO expresses his doubt whether they understand the essence of budgetary policy which relies on taking goods from some people in order to give them to the others. At this stage, at the onset of the EM, the pitch is slightly rising and speech is starting to gain more rhythmicity due to regular changes in intensity and pitch frequency. The speaker's engagement is gradually intensifying. He is making several beat gestures with both hands. These are short, energetic movements initially performed just above the rostrum, and then higher and higher (Figures 13, 14). The final beat in this part of the utterance is executed at the upper chest level. This part of the speech can be described as a crescendo (Figure 14). Then, DOSO directly addresses one of the members of parliament. During this part of the speech, he only makes two gestures. One pointing straight at the MP to whom the speech refers. The stroke phase in the described pointing gesture is executed at the head level. During the gesture, DOSO turns toward the MP and looks in their direction. The second gesture is a depictive (referential) gesture (Müller, 2024) embodying aspects of DOSO's spoken utterance. This part of the speech can be characterized as mezzoforte. Suddenly, an affective peak in the speech occurs, lasting approximately 11 s, during which 12 gesture phrases are executed. Most of them are quick, performed with both

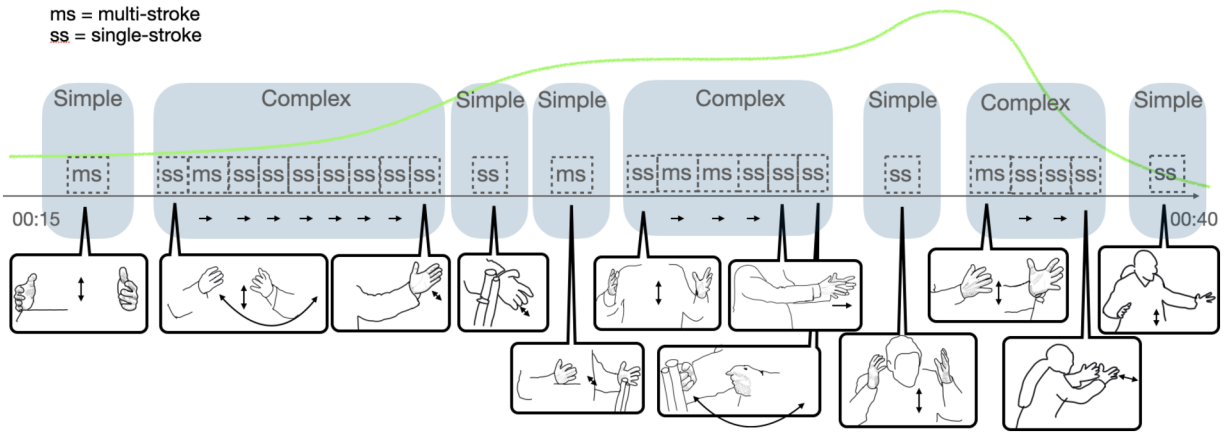


FIGURE 13
DOSO, the complexity of gestural performance.

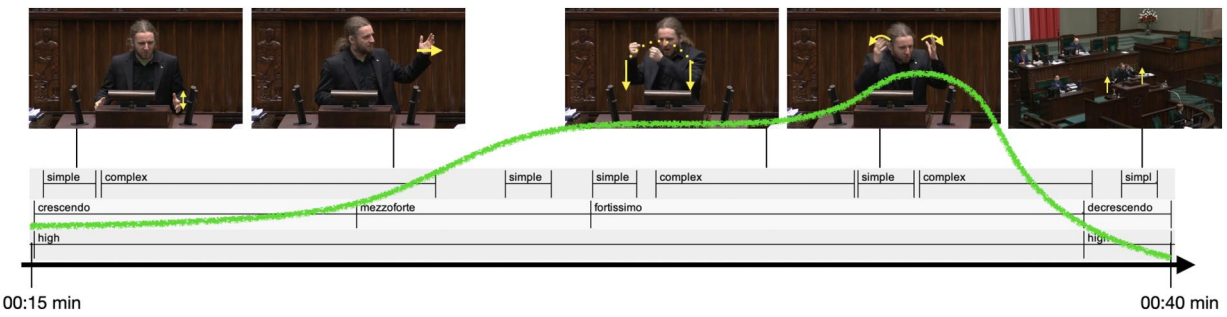


FIGURE 14
DOSO, vitality contour, musical terms, and the visibility of gestures in the image space as the EM unfolds. Recordings from parliamentary debates are sourced from System Informacyjny Sejmu.

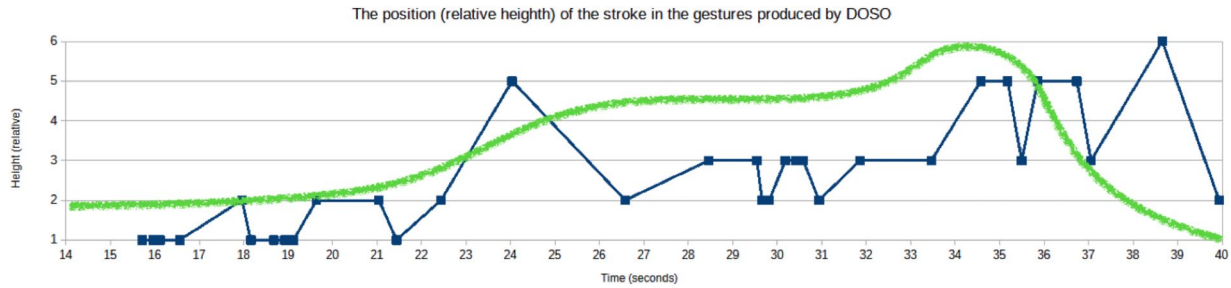


FIGURE 15
Vertical coordinate of the position of stroke realization (where "1" is the height of the abdomen and "6" is above the head).

hands, with a huge energy load in the beats, and high up in the gesture space.

Approximately halfway through this part of the speech, DOSO makes another depictive (referential) gesture. It enacts the action of giving and taking. Both hands performing the gesture are straightened at the elbow and mimic the transfer of an object from the left side of the speaker to the right side (Figure 14). This fortissimo part of the speech can be relatively easily identified on the basis of prosody itself. Pitch frequency remains on a distinctively high level, intensity is fluctuating, as well as the vocalic nPVI, making the rhythmic structure of this section even more prominent. Although clearly noticeable in earlier sections, speech-gesture synchrony is extremely vivid in the fortissimo section.

In the final part of the speech, described as a decrescendo, DOSO only performs beat gestures in the zone just above the rostrum (Figures 13, 14). His voice is gradually falling down to his baseline pitch frequency, pitch range is narrowing down. While intensity is not decreasing extremely, the voice may be perceived as slightly calming down and more stable.

4.5.2.1 Gestural complexity

In DOSO's speech, the dynamics of the EM also show in the complexity and visibility of the gestural performance—although the kind of complexity in which it unfolds is different from KAGE. DOSO is a very engaged gesturer, and he has very little time to speak. In fact, he gestures basically all the time along the unfolding of the Expressive Movement (approximately 25 s long). During this time, he produces a series of eight gesture units (Figure 13). The beginning of the EM evolves in a succession of a simple and a complex gesture unit, with the increasing affective dynamics (see the green vitality contour in Figure 13) we see a succession of five gesture units, the affective peak (fortissimo) unfolding as a sequence of complex, simple, complex units. The moment of high affective engagement terminates with a decrescendo and with the ending of complex and a simple gesture unit (Figures 13, 14). Most of the gestures realized within EM are single-stroke gestures (18 single-stroke gestures and 6 multistroke gestures).

In contrast to KAGE's performance, DOSO's EM does not show an increase in internal complexity. This indicates that there is not a simple match of high affective stance and complexity of gestural performance. This counters simplistic correlations of affective involvement with gestural performance, following the rule: the higher the affectivity, the higher the complexity of gestures. The multimodal orchestration of EMs is more variable than this.

4.5.2.2 Visibility of gestures and audiovisual orchestration

On the other hand, when it comes to the visibility of gestures, we see that, similar to KAGE's moment of high affectivity, DOSO moves his gestures increasingly higher up in the gesture space, thus moving them into the focus of joint visual attention (Figure 14). So, during the crescendo and mezzo-forte moments, most of the gestures are produced just over the rostrum, while in the peak phase, they are predominantly executed at the height of the upper chest area or higher. In short, moments of high affective dynamics appear to go along with a tendency to raise hands—also in DOSO's speech. Figure 15 shows how the position of the stroke in the gesture space changes along with the vitality contour of the expressive movement.

The green vitality contour (superimposed manually) illustrates how also in DOSO's EM different aspects of movement dynamics work together in forming this moment of high affective engagement (Figures 14, 15). We see a heightened visibility of the gestures and a change in movement dynamics of gestural and speech performance in particular in the affective peak phase.

We also see similar camera work in the Sejm as in the Bundestag recordings of the speeches. While most of the time, the speaker is placed in the center of the image space (with a medium shot), toward the end of the EM the camera perspective changes to a wide shot, reducing the size of the speaker for the spectators and hence the perceivable expressivity of the gestural performance. Altogether, the move *in camera*-perspective orchestrates the decrescendo phase of the EM.

4.5.3 Summary

Micro-analyses of the gestural performance and its audiovisual staging have shown that moments of high affective engagement go along with an increased complexity in KAGE's case and heightened visibility of gestural movements in both EMs. Gestures in high affective areas are performed in the upper part of the gesture space, which makes them more visible to the viewers on the screens and in the audience and increases their relevance as objects of joint attention. In areas with high affective engagement, gestures thus appear to become foci of joint visible attention. In both recordings, the placement of the gestures in the gesture space and the size of the gestures in relation to the image space is modulated by a change between medium shot and wide shot toward the end of the affective peak thus reducing visibility and hence the perceivable expressivity of the speaker's gesturing, but also bringing the audience within the parliament into the 'image' and the attention of the spectators.

It is in this sense that audiovisual and verbal multimodality are intertwined and it is this integrated orchestration that modulates the affective perception of the spectators in their process of watching audiovisual documentations of political speeches.

5 Summary and discussion

The media-aesthetic framework proposed in this paper offers a perspective on multimodal affective stance-taking in specific media ecologies. Exemplified by analyses of parliamentary discourse, it addressed in particular the intertwining of audiovisual multimodality and the multimodality of speaking. It advocates a phenomenological approach to affective dynamics of audiovisual composition and multimodal debates because it is interested in how spectators are moved by the orchestration of televised debates. In order to do this, the concepts of expressive movement and vitality contour served as theoretical and methodological references.

5.1 EM and vitality contours as keys to analyzing affectivity as stance

Conceived as multidimensional experiential gestalts, multimodal EMs allow affectivity to be reconstructed as unfolding movement dynamics and offer a way to address the mobilization of affective stance as modulation of the perception of the viewers.

To illustrate this approach, a perception-based analysis of EM (movement dynamics as affective qualities) was combined with an instrumental analysis of speech prosody. This means two independent empirical approaches to identifying EMs as dynamic gestalts were applied. Instrumental measures of EMs did reveal similar characteristics of affective contours as in the subjective analysis of EMs.

In a micro-analysis of the temporal unfolding of the gestural performances within the EM, two aspects of the multimodal orchestration of areas with high affective engagement received particular analytic attention: the complexity and visibility of the gestural performance. These aspects were chosen because they appeared as pertinent characteristics of moments with a high level of affective engagement.

5.2 Multimodal orchestration of stance

The analysis of the audiovisual staging of the debates shows that, for example, in KAGE's speech, the EMs that the viewers of parliamentary debates perceive on their screens are not only the multimodal enactments of a speaker but include audience interventions or the absence of them. The highly affective stance of KAGE reacts to interventions and shows that affective stance even in such a rather monologic context as a parliamentary speech is a joint process of affective dynamics within the parliament, which then is framed in a particular manner for the spectators (shift from medium to wide shot).

Here, KAGE formulates a complex stance of not only 'being against' but 'being ready to fight against' the actions of the right margins. She is performing a multimodal speech act of accusation by verbally accusing and intensely and repetitively pointing with a stretched index finger in the center of the gesture space while also gazing at the ones being accused. The peak ends in a change of camera perspective: concurrently the members of the parliament are shown in the shot. The mise-en-scene positions them at the top center showing their faces and their applause modulating a sense of interactive support and wideness, while the accused ones are visible on the lower side of the image space only from the back. The unfolding of the complex stance is therefore grounded affectively. It is exactly this basic form of human understanding, the immediate affective experience that is understood as stance.

5.3 Comparison of the two speeches

In the German recordings, these interactions with the audience are clearly audible, while this is different in the Polish material. In the case of the Polish parliamentary speeches, a lesser degree of interactivity between the speakers and the audience is visible and hearable. As mentioned above, this may be due to a few factors, starting from the strictly technical ones (e.g., microphones intentionally set to limit the hearability of the voices from the hall), or probably also to cultural ones. While during the speech of DOSO a few voices can be heard from the hall, they do not seem to influence his speech in an obvious way. On the other hand, one cannot say that there is no emotional tension between many of the listeners and the speaker, who is often perceived as radical and controversial.

The televised debates appear to differ in their multimodal orchestration of affectivity: whereas DOSO's speech is shown as unfolding its affectivity, especially on the level of the speaker's performance, KAGE's speech is staged as a debate and includes audience interventions. Similar staging forms are observed in the material from the German Bundestag (cf. Kindler-Mathôt et al., 2025) but not in the Polish data, where noticeable interactions are very limited.

Further research will show whether this staging of the speeches as more monologic or more dialogic results from different cultural standards and/or whether these are part of a more general understanding of how parliamentary debates are conceived in these two cultures. Political science studies show that, for example, the architecture of parliaments can be an expression of a specific understanding of parliamentary work (Minkenberg, 2020). Clearly, the seating arrangements prefigure specific kinds of interactions.

What we see here is how, even in such 'neutral' media formats, an orchestration of political processes is realized. Affective dynamics such as rising tension, beating rhythms, forcefulness, or softness characterize them and shape the embodied perception of the spectators. In both speeches, affective stances are realized as dynamically unfolding in expressive movements. Such observations underline the relevance to studying the audiovisual portrayal of democratic processes and representation (Kindler-Mathôt et al., 2025).

6 Conclusion

This article has presented a media-aesthetic approach to analyzing the multimodality of speech in audiovisual media. We have argued that the empirical analysis of multimodal speech data appearing in audiovisual media requires a reflection and methodological consideration of their specific media ecologies. By starting from the physically perceptible unfolding of affective multidimensional forms

of experience, so-called EMs and vitality contours, and not from a specific linguistic phenomenon, we have suggested a significant change of perspective. In accordance with Goodwin et al.'s (2012) notion of Emotion as Stance, we take the unfolding of affectivity as the anchor point to describe stance-taking not primarily on the level of single utterances nor only on the level of the spoken word but as emerging and unfolding affectivity orchestrated by both the multimodality of speaking (gesture and speech) and the audiovisual multimodality. Affective stance is not primarily bound to a specific gesture or a specific verbal expression; rather, it unfolds as a dynamic contour that emerges as a holistic gestalt from the multimodal performance of a speaker and the audiovisual staging of the speech. When we research affective stance-taking in audiovisual broadcasts of political speeches, we thus need to consider the expressive character of audiovisual compositions. It is the expressivity of the audiovisual staging that engages with the embodied perception of the viewers as an aesthetic affective experience. For the viewers of the broadcast speeches, the perception of the multimodal performance of speaking is framed by audiovisual multimodality. Against a media-aesthetic approach, virtual interaction is as embodied as face-to-face interaction. It is the multidimensional experiential gestalt of the televised speech and gesture performance that touches the viewers. In short, taking an affective stance in the context of political debates in the media must be considered as feeling an affective stance in the first place.

We believe that this media-aesthetic perspective may contribute to a better understanding of the power of audiovisual formats such as TikTok, YouTube, or Instagram Reels in the political arena (cf. Greifenstein et al., 2018; Kindler-Mathôt, 2025; Kindler-Mathôt et al., 2025; Müller, in press; Müller and Kappelhoff, 2018). It can provide a valid starting point for future comparative research of multimodal stance-taking, be it across cultures, discourse genres, or different forms of political debates.

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

CM: Supervision, Writing – original draft, Writing – review & editing. MK: Supervision, Writing – original draft, Writing – review & editing. CK-M: Methodology, Visualization, Writing – original draft, Writing – review & editing. KK: Data curation, Software, Supervision, Writing – original draft, Writing – review & editing. EJ-N: Supervision, Writing – original draft, Writing – review & editing. JJ: Visualization, Writing – original draft, Writing – review & editing. KP: Methodology, Writing – original draft, Writing – review & editing. BS-S: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The authors are grateful to all the previous members of the project group for their input. The study was carried out in the German-Polish MuMo Stance project funded by the Polish National Science Centre (UMO-2018/31/G/HS2/03633) and Deutsche Forschungsgemeinschaft (MU 1587/3-1) (cf. note 2285). The German team is grateful for the funds received by the DFG in the context of the Beethoven funding line.

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

Abelin, Å., and Allwood, J. (2000). Cross linguistic interpretation of emotional prosody. *ISCA Tutor. Res. Workshop Speech and Emot.*, 1–4.

Avdan, N. (2017). Collaborative competition: stance-taking and positioning in the European Parliament. Avdan - Linköping: Linköping University Electronic Press.

Biber, D., and Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text Interdisc. J. Study Discour.* 9, 93–124. doi: 10.1515/text.1.1989.9.1.93

Boersma, P., and Weenink, D. (1992–2024). Praat:doing phonetics by computer[computer program]. Version 6.2. Available at: https://www.fon.hum.uva.nl/praat/ (Accessed October 18, 2024).

Bressem, J., and Müller, C. (2014). "The family of away gestures: negation, refusal, and negative assessment" in *Body—Language—Communication. Handbücher zur Sprach-und Kommunikationswissenschaft: 38.2*. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin/Boston: De Gruyter Mouton), 1592–1604.

Bressem, J., and Müller, C. (2017). The "negative-assessment-construction" – a multimodal pattern based on a recurrent gesture? *Linguistics Vanguard*. 3, 1–9. doi: 10.1515/lingvan-2016-0053

Breyer, T., Buchholz, M. B., Hamburger, A., Pfänder, S., and Schumann, E. (2017). Resonanz – Rhythmus – Synchronisierung: Interaktionen in Alltag. Bielefeld: Therapie und Kunst. transcript.

Dancygier, B. (2012). "Negation, stance verbs, and intersubjectivity" in Viewpoint in language. eds. B. Dancygier and E. Sweetser. *1st* ed (Cambridge University Press), 69–94.

Dancygier, B., and Vandelanotte, L. (2017). "Internet memes as multimodal constructions" Cognitive Linguistics. vol. 28, 565–598. doi: 10.1515/cog-2017-0074

Debras, C., and Cienki, A. (2012). "Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking," Amsterdam, Netherlands: International conference on privacy, security, risk and trust and 2012 International Confernece on Social Computing. 932–937.

Du Bois, J. W. (2004). Searching for intersubjectivity: 'too' and 'either' In Stance Alignment. 25th Conference of the International Computer Archive of Modern and Medieval English (ICAME), University of Verona.

Du Bois, J. W. (2007). "The stance triangle" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam: Benjamins), 139–182.

Ekman, P. (1972). "Bewegungen mit kodierter Bedeutung: Gestische Embleme" in *Emotions in the Human Face*. New York: Pergamon.

Ekman, P. (1992). Facial expressions of emotion: an old controversy and new findings. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 63–69. doi: 10.1098/rstb.1992.0008

Englebretson, R. (2007). "Stancetaking in discourse. Subjectivity, evaluation, interaction" in Stancetaking in discourse: an introduction. ed. R. Englebretson (Amsterdam: John Benjamins), 1–25.

Feyaerts, K., Brône, G., and Oben, B. (2017). "Multimodality in interaction" in The Cambridge handbook of cognitive linguistics. ed. B. Dancygier (Cambridge: Cambridge University Press), 135–156.

Freeman, V. (2015a). Prosodic features of stance acts. *J. Acoust. Soc. Am.* 138:1838. doi: 10.1121/1.4933849

Freeman, V. (2015b). The phonetics of stance-taking. Available at: https://www.academia.edu/17373948/The_phonetics_of_stance_taking (Accessed July 19, 2024).

Fuchs, T. (2017). "Intercorporeality: emerging socialities in interaction" in Intercorporeality and interaffectivity. eds. C. Meyer, J. Streeck and J. S. Jordan (Oxford: Oxford University Press), 3–25.

Gibbon, D. (2005). "Prerequisites for a multimodal semantics of gesture and prosody" In: *Proc. International Workshop on Computational Semantics (IWCS)*. ed. H. Bunt, 6.

Gibbon, D. (2011). Modelling gesture as speech: a linguistic approach. *Poznań Stud. Contemp. Linguist.* 47, 470–508. doi: 10.2478/psicl-2011-0026

Gibbon, D. (2013). TGA: a web tool for time group analysis. In Proc. of the Tools and Resources for the Analysis of Speech Prosody (TRASP). 66–69.

Gibbon, D., Klessa, K., and Bachan, J. (2014). Duration and speed of speech events: a selection of methods. *Lingua Posnaniensis* 56, 59–83. doi: 10.2478/linpo-2014-0004

Goodwin, C. (2017). Co-operative action: Cambridge University Press.

Goodwin, M., Cekaite, A., and Goodwin, C. (2012). "Emotion as stance" in Emotion in interaction. eds. M.-L. Sorjonen and A. Perakyla (Oxford: Oxford University Press), 16–41.

Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis" in Laboratory phonology 7. eds. C. Gussenhoven and N. Warner (Berlin/New York: Mouton de Gruyter), 515–546.

Greifenstein, S. (2020). Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy. Berlin/Boston: Walter de Gruyter.

Greifenstein, S., Horst, D., Scherer, T., Schmitt, C., Kappelhoff, H., and Müller, C. (2018). "Cinematic metaphor in perspective" in *Reflections on a transdisciplinary framework*. (Berlin/Boston: De Gruyter Mouton).

Haddington, P. (2006). The organization of gaze and assessments as resources for stance taking. Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies, 26, 281–328. doi: 10.1515/TEXT.2006.012

Horst, D., Boll, F., Schmitt, C., and Müller, C. (2014). "Gesture as interactive expressive movement: inter-affectivity in face-to-face communication" in Body—language—communication. Handbücher zur Sprach-und Kommunikationswissenschaft. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *38.2* (Berlin/Boston: De Gruyter Mouton), 2112–2125.

Kappelhoff, H. (2004). Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit (13774895). Berlin: Vorwerk 8.

Kappelhoff, H. (2008). Realismus: Das Kino und die Politik des Ästhetischen. Vorwerk 8. Berlin: Vorwerk 8.

Kappelhoff, H. (2015). The politics and poetics of cinematic realism (BV043712222). Columbia University Press. Available at: http://www.degruyter.com/doi/book/10.7312/kapp17072 (Accessed July 19, 2024).

Kappelhoff, H. (2018a). "Frontlines of community" in Hollywood between war and democracy. ed. D. Hendricks (Berlin/Boston: Walter de Gruyter).

Kappelhoff, H. (2018b). Kognition und Reflexion. Zur Theorie filmischen Denkens. Berlin/Boston: Walter de Gruyter.

Kappelhoff, H., Bakels, J.-H., Chung, H.-J., Gaertner, D., Greifenstein, S., Grotkopp, M., et al. (2016). eMAEX – Ansätze und Potentiale einer systematisierten Methode zur Untersuchung filmischer Ausdrucksqualitäten. Available at: https://www.empirische-medienaesthetik.fu-berlin.de/emaex-system/emaex_kurzversion/entwicklung_emaex/index.html (Accessed October 18, 2024).

Kappelhoff, H., and Müller, C. (2011). Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and feature film. *Metaphor Soc. World* 1, 121–153. doi: 10.1075/msw.1.2.02kap

Karpiński, M., and Klessa, K. (2018). Methods, tools and techniques for multimodal analysis of accommodation in intercultural communication. *Computat. Method Sci. Technol.* 24, 29–41. doi: 10.12921/CMST.2018.0000006

Karpiński, M., and Klessa, K. (2021). Linguist in the field: a practical guide to speech data collection, processing, and management. *1st* Edn. Poznań: Wydawnictwo Rys Tomasz Paluszyński.

Karpiński, M., Klessa, K., and Czoska, A. (2014). Local and global convergence in the temporal domain in polish task-oriented dialogue. *Speech Prosody* 2014, 738–742. doi: 10.21437/SpeechProsody.2014-136

Kendon, A. (1972). Review of kinesics and context. *Am. J. Psychol.* 85, 441–456. doi: 10.2307/1420845

Kendon, A. (2004). Gesture: visible action as utterance. *1st* Edn. Cambridge: Cambridge University Press.

Kindler-Mathôt, C. (2025). "Erfahrungsraum Interaktion. Die audiovisuelle Modulation eines Gesprächs als Zuschauererfahrung" in Un/Reale Interaktionsräume. Formen sozialer Ordnung im Spektrum medienspezifischer Interaktion. transcript. eds. C. Kindler-Mathôt, D. Leblebici, G. Marinsalta, T. Rückwart and A. Zaglyadnova. Bielefeld, transcript.

Kindler-Mathôt, C., Müller, C., Junge, J. K., and Papadopoulou, K. (2025). "Sprechen in audiovisuellen Medien: Systematik für eine medienästhetische Analyse multimodaler Kommunikation" in Sprache entgrenzen. Beiträge zu einer kulturwissenschaftlichen Linguistik. transcript. eds. R. T. Vallentin and D. Horst. Bielefeld, transcript.

Klessa, K. (2016). Speech annotation mining with annotation pro plugins. Poznań: Wydawnictwo Rys.

Klessa, K., and Gibbon, D. (2014). Annotation pro + TGA: automation of speech timing analysis. In LREC 2014, Ninth International Conference on Language Resources and Evaluation. Reykjavik, Iceland, European Language Resources Association (ELRA).

Klessa, K., Karpiński, M., and Wagner, A. (2013). Annotation pro – a new software tool for annotation of linguistic and paralinguistic features. In D. Hirst and B. Bigi (eds.), Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP) Workshop (51–54).

Ladewig, S. H. (2014). "Recurrent gestures" in Body – language – communication. An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem, vol. *38.2* (Berlin/Boston: De Gruyter Mouton), 1558–1575.

Ladewig, S. H. (2024). "Recurrent gestures: cultural, individual, and linguistic dimensions of meaning-making" in The Cambridge handbook of gesture studies. ed. A. Cienki. *1st* ed (Cambridge: Cambridge University Press), 32–55.

McNeill, D. (1992). Hand and mind: what gestures reveal about thought. Chicago: University of Chicago Press.

Merleau-Ponty, M. (2005). Phenomenology of perception. London: Routledge.

Minkenberg, M. (2020). Macht und Architektur: Hauptstadtbau, Demokratie und die Politik des Raumes. Springer VS.

Müller, C. (2008). Metaphors dead and alive, sleeping and waking. A dynamic view (15322983). Chicago: University of Chicago Press.

Müller, C. (2024). "A toolbox for methods of gesture analysis" in The Cambridge handbook of gesture studies. ed. A. Cienki (Cambridge: Cambridge University Press), 182–216.

Müller, C. (2025). "Sprache-Sprechen, Hören, Sehen, Fühlen. Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien" in Sprache entgrenzen. Beiträge zu einer kulturwissenschaftlichen Linguistik (1. Auflage). transcript. eds. R. T. Vallentin and D. Horst. Bielefeld, transcript.

Müller, C., and Kappelhoff, H. (2018). Cinematic metaphor. Experience – affectivity – temporality. Berlin/Boston: De Gruyter Mouton.

Müller, C., and Speckmann, G. (2002). "DeSignis 3" in Gestos con una valoración negativa en la conversación cubana. eds. L. E. Chauvel and M. Rector (Buenos Aires: Gedisa), 91–103.

Müller, C., and Tag, S. (2010). The dynamics of metaphor. Foregrounding and activation of Metaphoricity in conversational interaction. *Cogn. Semiot.* 10, 85–120. doi: 10.3726/81610_85

Ochs, E., and Schieffelin, B. (1989). Language has a heart. *Text* 9, 7–25.

Plessner, H. (1982). "Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtsein des anderen Ichs [in cooperation with Buytendijk, Frederik J. J.]" in Helmuth Plessner. Gesammelte Schriften: 7: Ausdruck und menschliche Natur. eds. G. Dux, O. Marquard and E. Ströker (Suhrkamp), 67–129.

Scherer, K. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Comm.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scherer, K., Clark-Polner, E., and Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *Int. J. Psychol.* 46, 401–435. doi: 10.1080/00207594.2011.626049

Sobchack, V. (1992). The address of the eye. A phenomenology of film experience. Princeton: Princeton University Press.

Stern, D. N. (1985). The interpersonal world of the infant. A view from psychoanalysis and developmental psychology. London: Karnac Books.

Stern, D. N. (2010). Forms of vitality. Exploring dynamic experience in psychology, the arts, psychotherapy and development. Oxford: Oxford University Press.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. Proceedings of the 5th Language Resources and Evaluation Conference, 1556–1559.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 December 2024
DOI 10.3389/fpsyg.2024.1477263

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Pilar Prieto,
Pompeu Fabra University, Spain
Ingrid Vilà-Giménez,
University of Girona, Spain, in collaboration with reviewer PP
Isabella Poggi,
Roma Tre University, Italy

*CORRESPONDENCE
Christoph Rühlemann
✉ chrisruehlemann@googlemail.com

RECEIVED 07 August 2024
ACCEPTED 04 December 2024
PUBLISHED 27 December 2024

CITATION
Rühlemann C and Trujillo J (2024) The effect of gesture expressivity on emotional resonance in storytelling interaction. *Front. Psychol.* 15:1477263. doi: 10.3389/fpsyg.2024.1477263

COPYRIGHT
© 2024 Rühlemann and Trujillo. 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 effect of gesture expressivity on emotional resonance in storytelling interaction

Christoph Rühlemann[1]* and James Trujillo[2]

[1]Deutsches Seminar - Germanistische Linguistik, University of Freiburg, Freiburg, Germany, [2]Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands

The key function of storytelling is a meeting of hearts: a resonance in the recipient(s) of the story narrator's emotion toward the story events. This paper focuses on the role of gestures in engendering emotional resonance in conversational storytelling. The paper asks three questions: Does story narrators' gesture expressivity increase from story onset to climax offset (RQ #1)? Does gesture expressivity predict specific EDA responses in story participants (RQ #2)? How important is the contribution of gesture expressivity to emotional resonance compared to the contribution of other predictors of resonance (RQ #3)? 53 conversational stories were annotated for a large number of variables including Protagonist, Recency, Group composition, Group size, Sentiment, and co-occurrence with quotation. The gestures in the stories were coded for gesture phases and gesture kinematics including Size, Force, Character view-point, Silence during gesture, Presence of hold phase, Co-articulation with other bodily organs, and Nucleus duration. The Gesture Expressivity Index (GEI) provides an average of these parameters. Resonating gestures were identified, i.e., gestures exhibiting concurrent specific EDA responses by two or more participants. The first statistical model, which addresses RQ #1, suggested that story narrators' gestures become more expressive from story onset to climax offset. The model constructed to adress RQ #2 suggested that increased gesture expressivity increases the probability of specific EDA responses. To address RQ #3 a Random Forest for emotional resonance as outcome variable and the seven GEI parameters as well as six more variables as predictors was constructed. All predictors were found to impact Eemotional resonance. Analysis of variable importance showed Group composition to be the most impactful predictor. Inspection of ICE plots clearly indicated combined effects of individual GEI parameters and other factors, including Group size and Group composition. This study shows that more expressive gestures are more likely to elicit physiological resonance between individuals, suggesting an important role for gestures in connecting people during conversational storytelling. Methodologically, this study opens up new avenues of multimodal corpus linguistic research by examining the interplay of emotion-related measurements and gesture at micro-analytic kinematic levels and using advanced machine-learning methods to deal with the inherent collinearity of multimodal variables.

KEYWORDS

gesture kinematics, emotional resonance, talk-in-interaction, storytelling, electrodermal activity

1 Introduction

Arguably the most fundamental distinction between us is that we all have our own body. Despite this divide—or because of it—we seek to pull others closer to us or be pulled closer to them as a way to facilitate a socioemotional connection with one another (Marsh et al., 2009, p. 334).

Closing the gap between bodies can be achieved in innumerable ways. In talk-in-interaction people regularly repeat one another's behavior, often without noticing (Tschacher et al., 2024; Koban et al., 2019): they recycle words others have just used, re-use their grammatical constructions, mimic their co-speech gestures, align with their body postures, adapt their breathing rhythms to their partner's, etc. This "cross-participant repetition of communicative behavior" (Rasenberg et al., 2020, p. 2) has been referred to under various denominations, including resonance (Tantucci and Wang, 2021), alignment (Fusaroli and Tylén, 2016; Garrod and Pickering, 2009; Rasenberg et al., 2020), interpersonal coordination (e.g., Duran and Fusaroli, 2017; Romero and Paxton, 2023; Schmidt and Richardson, 2008; van Ulzen et al., 2008; Konvalinka et al., 2023), accommodation (Giles et al., 1991), coupling (Goldstein et al., 2015), and synchronization (Koban et al., 2019; Mogan et al., 2017).

Resonance is theorized to be a ubiquitous interpersonal process (Palumbo et al., 2016); its ubiquity likely has its roots in the fundamental functions it serves in interaction. Resonance promotes social bonding (Marsh et al., 2009). It may be cognitively desirable as being in synch with others conserves computational resources by merging self- and other-representations and is thus less costly than being alone (Koban et al., 2019). For example, measuring neural activity during synchronous speech using fMRI, Jasmin et al. (2016) found that "detecting synchrony leads to a change in the perceptual consequences of one's own actions: they are processed as though they were other-, rather than self-produced." Resonance may also have a role in establishing common ground (Brone and Zima, 2014), increasing subsequent cooperation and affiliation (Wiltermuth and Heath, 2009), and supporting group cohesion (Jasmin et al., 2016). Even when not explicitly interacting, people tend to passively synchronize with one another at different levels. For example, people in the same room exhibit spontaneous synchrony of non-communicative bodily movements, whether simply sitting in view of one another (Koul et al., 2023), or sitting and rocking in a rocking chair (Richardson et al., 2007).

But resonance is not only a behavioral phenomenon. Resonance also plays out on the level of Interpersonal autonomic physiology (IAP), defined as "the relationship between people's physiological dynamics, as indexed by continuous measures of the autonomic nervous system (ANS)" (Palumbo et al., 2016, p. 99). For example, infants and mothers have been shown to synchronize their respiratory kinematics during phases of increased infant attention as indexed by decelerated heart beat (McFarland et al., 2020). Mothers' facial skin temperature was found to align with their infant's temperature when they watched their child participate in a series of play and stress phases through a one-way mirror (Ebisch et al., 2012). In married partners, respiratory sinus arrhythmia (RSA), the heart rate variation that occurs during the breathing cycle, was found to positively correlate with self-reported marital conflict (Gates et al., 2015). Passive listeners in public performances of classical music exhibited synchrony on a number of physiological measures (Tschacher et al., 2024), including Electrodermal Activity (EDA), a measure that is particularly indicative of affect and emotion.

This capacity of EDA, to index emotional arousal, makes it a valuable tool for research on resonance in *storytelling*. For storytelling is claimed to be driven by emotion. Its key function is a meeting of hearts: a resonance in the recipient(s) of the story narrator's emotion toward the story events (cf. Stivers, 2008). Insights into how emotional resonance is achieved in storytelling are beginning to flow from Physiological Interaction Research (e.g., Peräkylä et al., 2015). This paper aims to contribute to this line of inquiry. Its focus is on the role of *gestures* in emotion expression and emotion resonance in storytelling. Gestures are a core feature of face-to-face language use (Holler and Levinson, 2019; Kendon, 2004; Özyürek, 2017; Vigliocco et al., 2014), forming part of a multimodal Gestalt (Holler and Levinson, 2019; Trujillo and Holler, 2023) and providing both semantic and pragmatic meaning to the utterance (Kendon, 2017; Özyürek, 2014). What is more, gestures are an important tool in a story narrator's toolkit for engendering resonance as gestures can also express emotions (e.g., Goodwin et al., 2012, p. 16; Selting, 2010, 2012; Couper-Kuhlen, 2012), and even influence memory and information uptake when paired with emotionally salient speech (Asalıoğlu and Göksun, 2023; Guilbert et al., 2021; Levy and Kelly, 2020). What is more, the performance of the same gesture can be varied along multiple kinematic dimensions such as speed, force, and size, thus creating an infinite number of expressive effects (cf. Dael et al., 2013). Such kinematic modulation of gesture has also been linked to the expression of different social intentions (Peeters et al., 2015; Trujillo et al., 2018), although its effect on emotional resonance is not yet known.

To illustrate the interplay of gestures by a story narrator and the participants' EDA responses, consider Extract 1 and Figure 1. The extract showcases a conversational storytelling by speaker A to two female story recipients, in which she expresses her distriss at frequently being misrecognized as a man. Her storytelling is accompanied by a large number of gestures; for space considerations only the three co-climax gestures are given below the simultaneous speech in Extract 1:

The story revolves around the protagonist, speaker A, needing to use the same public bathroom twice and each time being denied entry to the women's bathroom by a cleaning lady. While the first time around the misunderstanding is resolved by the cleaning woman recognizing and admitting her mistake (line 28), speaker A is again denied entry a week later by another cleaning lady. This time the misunderstanding is only resolved when the cleaning lady *from the week before* (line 41) intervenes on A's behalf by loudly proclaiming ↑*!THAT'S! A WOMAN (.) [!THAT'S! A WOMAN*↑ (line 44). During this constructed dialog, A performs a dramatic character-viewpoint gesture twice: a large X with both arms by which the cleaning lady indicates that denying A entrance to the women's bathroom is wrong (cf. embedded figures 2, 3). In line 45, A returns to her own character-viewpoint to depict her ensuing frustration and exhaustion: after she introduces constructed dialog with I'm like she silently drops her hands from the arm rest and pulls the corners of her mouth (cf. embedded figure 4).

The story recipients display sympathy with A: speaker B utters *£°aw°£* in a low smiley voice already in pre-climax position (line 42), while speaker C responds with a lengthened = *holy [sh:]i:[:t* in post-climax position (line 50).

```
A:          =[like] no::rmally how it happens
            so it's happened <in the> Hauptbahnhof a lot,@
            (0.804)+
A:          because like tho:s:e
            (0.480)
A:          I don't know if you have ever used that bathroom,
            in the train station¿=
C:          =[no]
B:          =[no]
A:          it's (.) ↑!it! is↑ kind o:f: you ha-
            it's not intuitive which one's the men's
            which one's the women's like you need to really
            like <read the si:[    gn    ]>
C:                       [oh the one] you have to pay to
            [enter¿ like one]
A           [     !yeah!    ]
C:          euro [      or something    ]°yeah yeah°
A:               [yeah =yeah =yeah =yeah]
A:          and like IT'S AN OPEN entrance on either end (.)
            I get it you go in the wrong end
            (0.211)
            u::m (.)
            I was going in one day and the cleaning woman there
            !stopp!ed me >she was like<
            (0.384)
A:          ~er (0.237) sorry this is the women's~
            and I turned around she goes (.)
            ~((v: gasps)) I'M SO SORRY >>go in go in go in go in<<~
            (0.331)
A:          and like a !week! later the same bathroom going in the:re
            (0.394)
A:          and an!other! woman !stop!ped me
            like she's coming out the stall and she's like
            ~((v: gasps)) WAIT (.) THIS IS THE WOMEN'S '~ (.)
            and I'm like
            (0.850)
A:          and I just !look!ed at her and
            >she was also< much much shorter than me so
            <I'm like looking down at her>
            and the ↑cleaning↑ woman is there
            from the week be↑fore↑ who had ↑already↑ done this=
B:          =£°aw°£=
A:          @=>and she went< #!out of her way! to come in >and be like<@
A_gest:     @r hand palm open sweeps outward & returns toward chest--->@

fig         #fig. 1
```

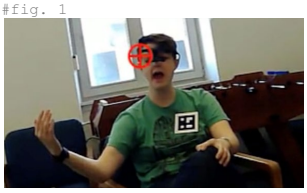


figure 1

```
A:          (0.280)  #*↑!THAT'S! A WOMAN≈ (.) !THAT'S! A WOMAN↑#
A_gest:     #both arms draw large X from chest to knee, 2x --->#

fig         *fig. 2                    ≈fig. 3
```

figure 2

figure 3

```
A:          I'm like*
A_gest:                *drops hands from arm rest, pulls corners of mouth*

fig          *fig. 4
```

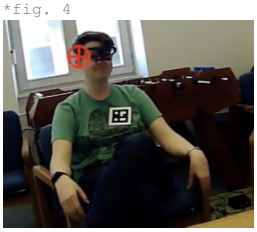


figure 4

```
B:          [((v: laughs))]=
C:          [((v: laughs))]
C:          =holy [ sh: ]i:[  :t does it happen ] often? [like]
A:                [jus:t]  [gonna (.) stand here]        [yeah]
```

The storytelling is performed with considerable gestural effort; altogether 25 gestures were annotated. Does the effort have any measurable effect on the recipients emotionally? Figure 1 depicts the EDA responses of all three participants during the telling.

As shown in Figure 1, in pre-climax positions, only the story narrator seems to experience changes in EDA, whereas the two recipients' EDA responses remain largely flat. This changes during the climax: not only are there huge increases in EDA in the story narrator A, but there is a clear simultaneous hike in EDA in recipient B, specifically during the third in-climax gesture; in recipient C, however, EDA does not change at all.

As far as this storytelling is concerned, the story narrator and recipient B seem to resonate emotionally: their EDA responses peak at the same time, roughly in synchrony with the story narrator's gestures.

The overarching goal in this paper is to examine whether the resonances that can be observed between participants in conversational storytelling interaction are an effect of the expressivity of the story narrator's gestures. Early pioneering found that different emotions are expressed through different parameters of bodily movement and posture (Wallbott, 1998), and that bodily movement can express the intensity of emotions (Ekman and Friesen, 1974). Here, we focus on expressivity of co-speech gestures, using movement features that have been shown to be important indicators of expressivity. Specifically, the current measure of gesture expressivity captures similar dynamic aspects of gesture as previous work interested in quantifying expressivity, for example in the gesture synthesis literature, that independently assess features such as spatial and temporal extent, power, and fluidity (Hartmann et al., 2005; Pelachaud, 2009). Visual annotation of these features has similarly been applied to analyzing expressivity of observed gestures (Chafai et al., 2006; Kipp et al., 2007). The quantification of gesture expressivity used here captures many of these same kinematic features (i.e., size, force, hold-phases, bodily co-articulation, nucleus duration), but also includes dialogically relevant non-kinematic features (i.e., character-viewpoint, silence during gesture) and aggregates them into a singular value. This allows a more straightforward assessment of the role of gesture expressivity that accounts for both kinematic parameters as well as characteristics related to the gesture's embedding in the ongoing multimodal narrative. Clearly, emotional resonance in conversational storytelling can result from factors and their interactions that are not part of or related to gesture performance as such but originate in the story narrator's verbal performance and/or the design of the storytelling situation. Therefore, while we focus on the potential effect of gesture expressivity on emotional resonance we will approach the effect multi-factorially, by considering a large number of potentially contributing factors besides gesture expressivity.

As a first step to understanding the role of gestures in emotional resonance of conversational storytelling, we also assessed whether gesture expressivity follows the same "climacto-telic" (Georgakopoulou, 1997; Labov, 1972) structure that has been described for the oral component of storytelling. Climacto-telic refers to the gradual build-up of "tension" (Longacre, 1983) to be released only at climax. Initial observations gained from small-scale empirical analyses suggest that story narrators advance-project the story climax using an orchestrated crescendo of expressive multimodal means

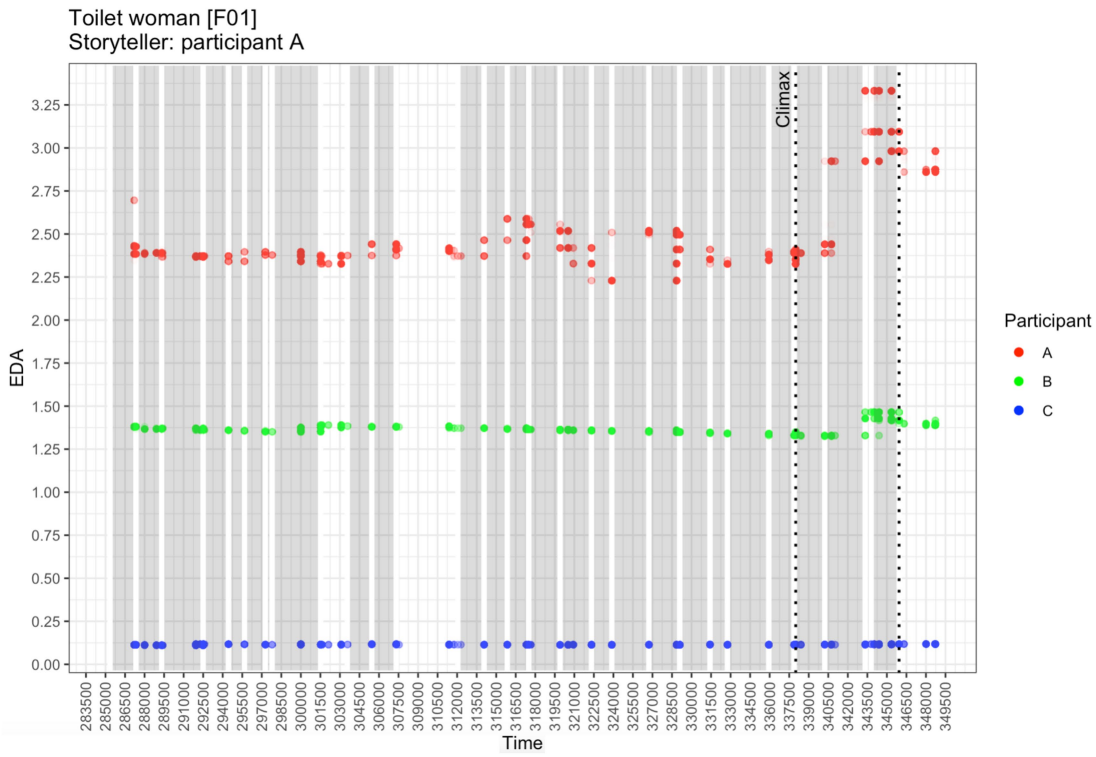


FIGURE 1
EDA responses in story "Toilet woman" by participant; gray rectangles demarcate the extensions of the story narrator's gestures. Colored dots represent observed EDA measurements (indicated along the y-axis). Time is given along the x-axis, in milliseconds from the beginning of the recording.

including constructed dialog, pitch, intensity, gaze alternation, and gestures (Mayes, 1990; Holt, 2007; Goodwin, 1984; Rühlemann, 2019; Rühlemann et al., 2019; Rühlemann, 2022).

Specifically, the paper asks three questions: Does story narrators' gesture expressivity increase from story onset to climax offset (RQ #1)? Does gesture expressivity predict specific EDA responses in story participants (RQ #2)? How important is the contribution of gesture expressivity to emotional resonance compared to the contribution of other predictors of resonance (RQ #3)?

2 Methods

2.1 The FreMIC corpus

The data underlying the analyses in this paper come from the Freiburg Multimodal Interaction Corpus (FreMIC; cf. Rühlemann and Ptak, 2023). FreMIC is a multimodal corpus of naturalistic conversation in English, which is, at the time of writing, still under construction. All conversations were annotated and transcribed in ELAN (Wittenburg et al., 2006). The transcriptions follow conversation-analytic conventions (e.g., Jefferson, 2004) to render verbal content and interactionally relevant details of sequencing (e.g., overlap, latching), temporal aspects (pauses, acceleration/deceleration), phonological aspects (e.g., intensity, pitch, stretching, truncation, voice quality), and laughter.

2.2 Participants

Fourty-one individual participants were recruited to contribute to one or more of the 38 recorded conversations (total run time 30 h). The participants were mainly students at Albert-Ludwigs-University Freiburg as well as their friends and relatives [17 male, 21 female, 3 diverse/NA; mean age = 26 years (SD = 5.7 years)]. Most participants' (n = 38) first language was English. All participants had normal or corrected to normal vision and hearing. Before the start of the recording, participants gave their informed consent about the use of the recorded data, stating their individual choices as to which of their data can be used and for what specific purposes. They received a compensation of €15 for their participation.

2.3 Procedure

Recordings were made in dyadic and triadic settings using one room camera and one centrally placed scene microphone. Seated in an F-formation (Kendon, 1973) participants were able to establish eye contact, hear each other clearly, and engage in nonverbal cues. Participants in dyads were seated vis-à-vis each other, with the room camera capturing both participants from the side. Participants in triads were seated in an equilateral triangle, with the room camera frontally capturing one of the participants and the other two from the side. The participants were told they were free to talk about anything for about 30–45 min until the recording would be stopped.

Participants wore Ergoneers eyetracking devices (Dikablis Glasses 3), which recorded the visual field of each participant plus the direction of participants' gazes. Participants wore also Empatica wrist watches, which recorded a wealth of psycho-physiological data, including Electrodermal Activity (EDA), the measurement of central interest in this study. The wrist watches' sampling frequency for EDA measurements is 4 Hz within a range of 0.01–100 μSiemens. Due to malfunction of the Empatica wrist watches, only nine recordings produced EDA data.

With the watches being placed at the wrists, EDA is measured in close proximity to the palms, where the highest concentration of eccrine sweat glands is found. The sweat produced by these glands is emotion-evoked (Dawson et al., 2000, p. 202) rather than thermo-regulatory (Bailey, 2017, p. 3; cf. also Scherer, 2005; Bradley and Lang, 2007), making palm-near EDA measurements a reliable indicator of emotional arousal (Peräkylä et al., 2015). Arousal is defined as the intensifying excitation of the sympathetic nervous system associated with emotion (Dael et al., 2013, p. 644; Peräkylä et al., 2015, p. 302). Heightened arousal results in increased EDA while emotional unaffectedness correlates with decreases in EDA. Being controlled by the sympathetic nervous system neither process can be influenced volitionally.

The focus in the present analysis is on phasic EDA representing "transient, wave-like changes which may be elicited by external stimuli or may be "spontaneous," i.e., elicited by internal events" (Lykken and Venables, 1971, p. 657). Given the overall aim to examine the possible effect of gesture expressivity on emotional resonance, the phases during which all participants' EDA is examined are the story narrator's gestures (plus an additional time window to account for response latency; cf. Section 2.4.4).[1]

2.4 Data pre-processing

2.4.1 Story selection

53 conversational storytellings were selected for this analysis. Given the scarcity of recordings with EDA measurements (cf. Section 2.3), the selection criteria for stories were relatively broad. The only must-have criteria were (i) anterior situation (the story events happened in the past rather than in the future or in an imaginary world, cf. Norrick, 2000) (ii) involving at least one a-then-b relation (i.e., the temporal sequencing of at least two narrative events; cf. Labov and Waletzky, 1967),[2] (iii) extension (all stories except one are longer than half a minute, thereby excluding so-called "small stories," cf. Bamberg, 2004), and (iv) the use of constructed dialog (also referred to as "quotes," "direct speech" or "enactments"; cf. Labov, 1972).

Story climaxes were identified as those story events that semantically matched the emotion expressed at story onset. For

1 EDA responsiveness varies substantially between subjects. This variability is mostly due to physiological conditions (such as the thickness of the cornea); rarer factors impacting on responsiveness include psychopathological conditions (Lykken and Venables, 1971; Dawson et al., 2000).

2 One story ("Pay more"), which is the shortest in the sample with 10.9 s, consists of exactly *one* a-then-b relation only: *[£I lit]erally had a call today with the health insurance£ £where they were like saying ~ oh we need to update your document (.) you'll [have to pay] more ~ and I was like ~ (↑I do not wanna pay more↑)£ ~ [((v: laughs))] ((v: laughs))*, where the two examples of direct quotes, indicated by "~," represent the two narrative events.

TABLE 1 Interrater agreement on gesture expressivity index (GEI) parameters (based on 227 gestures, or 22%, out of 1,021 gestures in total); the GEI parameters *Hold phase* (*HO*), *MA* (Co-articulation with other bodily organs), and Nucleus duration (ND) were calculated directly from their annotations in ELAN.

GEI parameter	Agreement %	Interrater reliability (G)
CV	94.27	0.89
FO	78.41	0.57
SL	97.36	0.95
SZ	86.78	0.74

example, the climax to a story billed as "sad" was the story's sad(dest) event, the fun(niest) event was coded as the climax in a "funny" story. Another identification criterion was the occurrence of direct speech; this criterion relies on the widely accepted notion that direct speech clusters at story climaxes (cf. Labov, 1972; Li, 1986; Mathis and Yule, 1994; Mayes, 1990; Norrick, 2000; Clift and Holt, 2007; Rühlemann, 2013). Another criterion used to identify a story's climax was "texturing" (Goodwin, 1984), that is, variations in the story narrator's paralinguistic prosody such as raised pitch and increased intensity. Finally, given that storytellings constitute activities centered essentially around stance and emotion, climaxes are characterized by recipients mirroring the story narrator's stance/emotion displayed earlier (implicitly or explicitly). Verbally, that mirroring is achieved through the use of tokens of *affiliation* displaying the recipient's stance, including, for example, assessments such as *wow* (Goodwin, 1986), head nods (Stivers, 2008) or laughter (Sacks, 1978).[3]

2.4.2 Story annotation

The stories were rated for a large number of variables that may potentially impact emotion arousal. These include (i) *Protagonist* (whether the story's protagonist is the story narrator or a non-present third person; cf. Ochs and Capps, 2001), (ii) *Recency* (whether the story events occurred far in the past or were occurring at or close to storytelling time), (iii) *Group_composition* (whether groups were all-female, all-male, or mixed), and (iv) *Group_size* (whether the storytelling setting was dyadic or triadic).

The variables *Protagonist* and *Recency* were considered potentially impactful for emotional resonance as they are components of *relevance*, a key dimension of emotion, as "in order for a particular object or event to elicit an emotion, that object or event needs to be […] relevant to the person in whom that emotion is elicited" (Wharton et al., 2021, p. 260). Indeed, emotions can be seen as "relevance detectors" (Scherer, 2005, p. 701). The factor *Group_size* is included somewhat tentatively based on the recent finding that response times are faster in triads than dyads (Holler et al., 2021) due to competition. While, obviously, competition in triads is not emotion *per se*, competition may nonetheless contribute to heightened emotion arousal. *Group_composition* was included as predictor to capture potential effects of gender. In Doherty (1997), for example, women were highly significantly more susceptible to emotion contagion than men (but see Eisenberg and Lennon's, 1983 large meta-study, in which females did not exhibit more empathy than males when physiological measures were used to index empathy).

The variable *Sentiment* was calculated to account for the emotional impact of individual words uttered in each interpausal unit (IPU). Sentiment analysis was performed on individual words using the Python package *Vader* (Hutto and Gilbert, 2014), providing the associated positive, negative, and composite (combined) sentiment scores for each transcribed word. The composite sentiment score was taken, which reflects the composite of both positive and negative scores for each given word. The mean composite score was calculated for each IPU, and each gesture (described below) was assigned the *Sentiment* score of the IPU with which it occurred.

2.4.3 Gesture annotation

1,021 gestures as well as their gesture phases (Kendon, 2004) were identified and annotated in ELAN by multiple raters. The gestures were further coded in ELAN for seven gesture-dynamic parameters: (i) Size (*SZ*; Dael et al., 2013), (ii) Force (*FO*; Dael et al., 2013), (iii) Character view-point (*CV*; McNeill, 1992), (iv) Silence during gesture (*SL*; Hsu et al., 2021, p. 1; Kendon, 2004, p. 147; Siddle, 1991, p. 247), (v) Presence of hold phase (*HO*; Beattie, 2016, p. 129; Gullberg and Holmquist, 2002), (vi) Co-articulation with other bodily organs (*MA*; Dael et al., 2013; Rühlemann, 2022) and (vii) Nucleus duration (*ND*; Kendon, 2004).

In judging gesture size (SZ), lateral and forward movements were distinguished. Size in lateral movements was coded based on McNeill's (1992) gesture space schema. A gesture was considered sizable if it crossed at least two major lines in the gesture space schema; e.g., from CENTER-CENTER to PERIPHERY, or from EXTREME PERIPHERY to CENTER. If the onset of a gesture was not at the "normal" rest position (i.e., in the speaker's lap or on the arm rest) but at some other point in the gesture space, that onset was taken as the starting point of the gesture's trajectory and the count of how many major boundaries it crossed started from there. For example, if the gesture's onset was in the right EXTREME PERIPHERY and moved back to PERIPHERY, one single major boundary is crossed and the movement was considered not sizable. Gesture size can also become expansive if the hands' and arms' orientation is away from the gesturer's body into the space in front of them, i.e., if they extend their hands toward the interlocutor. Forward gestures were coded sizable only if the speaker extended her arms beyond a 45° degree angle.

Gestures were coded forceful (FO) based on the requirement that the movement requires muscular effort. To gage whether muscular effort was involved, annotators physically reenacted the gesture. Also, a diagnostic of a forceful gesture is the whiplash effect, i.e., when the hand slightly bounces back from the gesture's endpoint. Gesture force undoubtedly enters into a number of interactions with other dynamic parameters. Clearly, the more sizable a gesture the more muscular effort it will involve. Also, extended holds (especially of sizable gestures) likely require muscular effort, as do gestures that are carried out fast, again especially if they are sizable. The fact that FO (force) saw the least interrater agreement (cf. Table 1) is therefore not surprising.

Gestures were coded as character-viewpoint gestures (CV) if they were carried out as if the gesturer slipped into the role of the character; alternatively, the gesture is performed from the gesturer's own perspective, as if observed by them (cf. McNeill, 1992; Beattie, 2016). While character-viewpoint gestures are often representational gestures, in most co-quote gestures the movements carried out by the gesturer were the character's movements not the gesturer's regardless of type of

3 For illustration, cf. the very engaged response tokens *£°aw°£* (line 43) and *holy sh:i::t* (line 48) surrounding the climax in excerpt (1).

gesture. The inclusion of CV among the (potentially) expressive gesture features is based on the observations that (i) character-viewpoint gestures are more effective for communicative purposes than observer viewpoint gestures (Beattie, 2016), (ii) they represent "demonstrations" rather than "descriptions" (Clark and Gerrig, 1990) allowing the storytelling recipients to immediately and immersively see and experience the displayed emotions without the observer's intermediary perspective separating the audience from them, and (iii), in our data, they overwhelmingly occur within direct quotation, a discursive practice that facilitates heightened multimodal activation in speakers (e.g., Blackwell et al., 2015; Stec et al., 2016; Soulaimani, 2018).

Silent gestures (SL), alternatively referred to as "speech-embedded non-verbal depictions" (Hsu et al., 2021), are gestures that communicate meaning "iconically, non-verbally, and without simultaneously co-occurring speech" (Hsu et al., 2021, p. 1). With the (default) verbal channel muted, the burden of information is completely shifted to bodily conduct (cf. Levinson and Holler, 2014, p. 1). This shift makes silent gestures particularly expressive: they are "foregrounded" and "exhibited" (Kendon, 2004, p. 147).

Actively attending to them is prerequisite for the recipient's understanding. Moreover, given that the occurrence of speech is expected, its absence will not only be noticeable but also emotionally relevant as the omission of an expected stimulus has been shown to increase EDA response (Siddle, 1991, p. 247).

Subsuming the presence of a hold phase (HO) under expressive gesture dynamics draws on the absence of movement, which is assumed to gain saliency considering the lack of progressivity manifested in the hold. Given the preference for progressivity (Stivers and Robinson, 2006; Schegloff, 1997), which we assume extends to a preference for progressivity in bodily conduct, the uninterrupted execution of a gesture can be seen as preferred, aligned with the default expectation of progressive movement, whereas the interrupted execution as occurring during a hold phase will be seen as disaligned with the default expectation of progressive movement and hence dispreferred. As a dispreferred, the gesture hold, just as a "hold" during speech, "will be examined for its import, for what understanding should be accorded it" (Schegloff, 1997, p. 15) and is therefore likely to raise attention and add to the saliency of the gesture. Gesture holds have also been linked to saliency in silent gesture paradigms (Trujillo et al., 2018), where longer hold-times are associated with better gesture recognition (Trujillo et al., 2020). Also, while the overwhelming majority of gestures are not gaze-fixated, i.e., not taken into the foveal vision, but still processed based on information drawn from the parafoveal or peripheral vision (cf. Beattie, 2016, p. 129), those gestures that contain a hold phase, i.e., a momentary cessation in the movement of the gesture, reliably attract higher levels of fixation (Gullberg and Holmquist, 2002). Beattie argues that "[d]uring 'holds', the movement of a gesture comes to a stop and thus the peripheral vision is no longer sufficient for obtaining information from that gesture, thus necessitating a degree of fixture" (Beattie, 2016, p. 131).

The annotations were implemented using a binary scale (yes/no) and aggregated in the Gesture Expressivity Index (GEI). The Index computes for each gesture an average value across all yes/no ratings; the Index values are stored in the variable *G_expressivity*, one of the key variables in the models. Note that this GEI value is thus a combination of kinematic salience values (4 features): SZ, FO, HO, and ND), multimodal coarticulation (two features, SL and MA), and one capturing a broader narrative embodiment (CV). There is, therefore, a somewhat heavier weighting of kinematic features in the calculation of GEI compared to other features. To date, there is no research indicating the actual weight of each of these features in predicting how "expressive" a gesture is perceived to be. Additionally, there was the possibility that character viewpoint gestures, which are defined more broadly (i.e., not based on more fine-grained movement parameters) would also be more kinematically salient. As a basic check, we conducted simple chi-square tests to assess whether there was evidence for the character viewpoint gestures being more likely to be larger, more forceful, or more likely to have a hold-phase than non-character-viewpoint gestures, and we did not find evidence for this to be the case (p-values from 0.635 to 0.921). For the purpose of this study, we therefore take the entire set of seven features as being relatively independent and equal in importance. The distribution of the GEI parameter codings is shown in Figure 2.

As shown in Figure 2, a third of all gestures had a hold phase (HO) and in roughly a quarter of them the duration of the nucleus was longer than the average nucleus in the respective story (ND); on the other hand, gestures were rarely coded as character viewpoint gestures (CV, 6%), gestures performed together with other bodily articulators (MA, 4%) and silent gestures (SL, 2%). Across all gestures, GEI ranged from 0 to 0.857, with a median value of 0.286. This indicates that most gestures (79%) showed two or fewer GEI features.

Interrater agreement for the coding of the GEI parameters was tested on c. 22% of all 1,021 gestures. Interrater agreement percentages ranged between 78% for Force (*FO*) and 97% for Silent gesture (*SL*). To calculate interrater reliability we used Holley and Guilford (1964) *G* index; the G values are reported in Table 1. We chose to use *G* as it provides a more robust measure of reliability than Cohen's Kappa when the rating distributions are skewed (Silveira and Siqueira, 2023; Xu and Lorber, 2014). In the case of our GEI parameters, there was a heavy skew toward "negative" ratings, meaning that most gestures were coded as *not* having a particular parameter. CV and SL ratings showed near perfect reliability, SZ showed substantial reliability, and FO showed moderate reliability. While the lower *G* index for FO indicates that gesture force is a more difficult parameter to reliably code, the lower *G* is likely also due to the skewed coding. Specifically, 78% of the reliability-coded gestures received a *no* coding from at least one rater, with a true negative rate of 56%.

The parameters HO (hold), SL (silent gesture), and multiple articulators (MA) were extracted from the ELAN gesture and gesture phase annotations and were therefore not tested for interrater reliability. For example, whether a gesture was performed together with other bodily articulators was read off the gesture descriptions that indicated all articulators used. For example, the last in-climax gesture in extract (1) has this description: *((112_m & f: both h rested on arm rest slightly drop, pulls corners of mouth))*. Here, the initials "m" and "f" refer to manual (hand) and face as the articulating organs.

Further, based on FreMIC's existing annotation of quotes (alternatively referred to as direct speech [e.g., Labov, 1972), constructed dialog (e.g., Tannen, 1986), or) and enactments (e.g., Holt, 2007)] and using a fuzzy assignment procedure which allowed for a durational "distance" of 1.5 s. of the respective start times, the gestures were examined for whether they co-occurred with a quote (variable *G_quote*). Instances of direct speech are likely to impact emotional resonance not only as they thrive in storytellings (Rühlemann, 2013; Stec et al., 2016) but also because they facilitate heightened activation in the speaker's vocal and bodily channels (e.g., Blackwell et al., 2015;

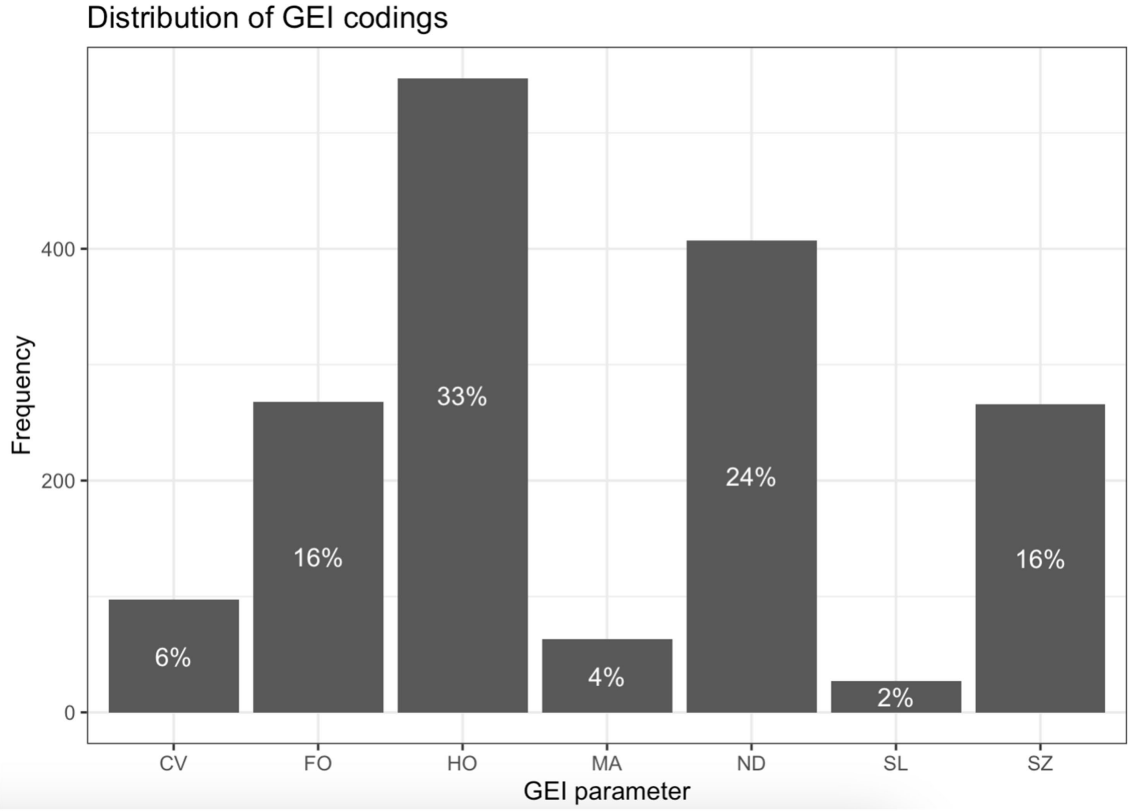


FIGURE 2
Frequency of Gesture Expressivity Index (GEI) features where coding was "yes". The seven GEI parameters are given along the x-axis. The height of the bars indicates the percentage out of all gestures where a given parameter was coded as having a particular feature.

Stec et al., 2016; Soulaimani, 2018) and because they are frequently mimicry (e.g., Mathis and Yule, 1994; Stenström et al., 2002, p. 112; Halliday and Matthiessen, 2004, p. 447), "a caricatured re-presentation" (Culpeper, 2011, p. 161) or "echo" of anterior discourse, "reflect[ing] the negative attitude of the echoer toward the echoed person" (Culpeper, 2011, p. 165).

The total number of stories the present analyses are based on 53 stories collected in nine recordings (total run time 7.55 h), with 1,021 gestures by the story narrators, and 14 distinct participants (Table 2).

The familiarity levels between the participants in the recordings were mixed throughout. While in some triadic recordings either siblings or romantic partners participated (with high familiarity), the third subject was always either a stranger (low familiarity) or an acquaintance (medium familiarity), while the dyadic conversations were invariably between friends or acquaintances. Applying a three-level distinction (low, mixed, or high familiarity) would have resulted in a single value, "mixed." We therefore decided not to use familiarity as a predictor in our models.

2.4.4 EDA pre-processing

To account for response latency (typically between 1 and 3 s; Dawson et al., 2000, p. 206), EDA responses were measured during the duration of the gesture as well as 1.5 s post-gesture. Further, EDA responses were classified as *specific* (i.e., as indexing a stimulus-related emotional response) if they were larger than 0.05 μSiemens. Finally, *resonating gestures* were identified, on the condition that they exhibited *concurrent* specific EDA responses by two or more participants. The result is a binary variable *EDA_G_resonance*, which represents the dependent variable in the Random Forest model (see below).

2.5 Statistical analysis

RQ #1 was addressed using a mixed-effects model. In order to handle the large variance in the number of gestures used in the stortyellings (the range is 4–91 gestures; mean = 21.6; median = 15.5), a relative positional measure *G_position_rel was* computed for each gesture in each story assigning as many equi-distanced values between 0 and 1 as there are gestures in the storytelling (e.g., the relative positions of 4 gestures are 0, 0.25, 0.5, 0.75, and 1). The fixed effects in the model were *G_expressivity* (as the response variable) and *G_position_rel* (the independent variable); the random variable was ID, a combination of participant and recording ID.

To adress RQ #2, a second linear mixed-effects regression model was constructed, with *EDA_specific_response_binary* as the dependent and *G_expressivity* as the independent variable, as well as *Participant* and *Recording* modeled as random effects. If there were not issues with model fit, we modeled random slopes, rather than random intercepts.

For the mixed models described for RQ #1 and #2, statistical significance was determined using a log-likelihood test, comparing the full model as described above against a null model having the same structure but without the main independent variable of interest. For RQ #1, the null model did not include *G_position_rel*, while for RQ #2, the null model did not include *G_expressivity*. We report conditional pseudo-R^2, as calculated by the *MuMIn* R package (Bartoń, 2009) as a measure of effect size.

RQ #3 requires a different statistical approach, as it specifically addresses the magnitude of the influence of *G_expressivity* and, respectively, the gesture dynamics that feed into it, on the outcome

TABLE 2 Sociodemographics of FreMIC participants to this study.

Sex	*N*	Mean age	Age range	SD	L1
Female	5	26	22–31	3.39	English: 4; other: 1
Male	8	31	24–49	9.08	English: 8
NA/Diverse	1	27	27	*NA*	English: 1

TABLE 3 Output of linear mixed-effects regression model on *G_expressivity* in storytellings.

Fixed effects	*β*	*SE*	*df*	*t*	*p*	*Sig*
(Intercept)	2.048e-01	6.484e-02	7.965 e-02	31.585	<0.001	***
G_position_rel	2.786e-02	1.585e-02	1.825 + 03	1.785	0.095	

Formula: G_expressivity ~ G_position_rel + (1 | ID).

TABLE 4 Output of linear mixed-effects regression model on G_expressivity in storytellings.

Fixed effects	*β*	*SE*	*z*	*p*	*Sig*
(Intercept)	−0.9489	0.4183	−2.269	0.0233	*
G_expressivity	1.4314	0.3378	4.237	2.26e-05	***

Formula: EDA_G_specific_responses_binary ~ G_expressivity + (1 | ID).

variable *EDA_G_resonance* relative to the magnitudes of other potentially impactful predictors, most of which can be assumed to be highly collinear. The method warranted by this type of research scenario is a Random Forest model. Random Forests are able to handle collinear features effectively, based on vertical sampling of variables (feature subsampling), horizontal sampling of data (bootstrapped sampling), and random decision tree splitting based on a single predictor at a time thus mitigating the influence of collinearity not only within individual trees but also across trees. Another advantage of Random Forests is that they indicate relative variable importances (cf. Tagliamonte and Baayen, 2012; Gries, 2021). The Random Forest built here comprises 1,500 trees (ntree = 1,500) and three randomly preselected predictors at each split (mtry = 3).

To investigate the effect on emotional resonance of potential interactions between predictors, Individual Conditional Expectation (ICE) plots (Goldstein et al., 2015) were inspected. These plots visualize how changes in a single predictor affect the predicted response of a model *for each individual instance*. Each line in an ICE plot represents the prediction for a single case of the response (here, gesture-related emotional resonance) as the predictors of interest change while keeping all other predictors constant. ICE plots are particularly useful for identifying interactions between the response and the predictors in the model.

3 Results

3.1 Does story narrators' gesture expressivity increase from story onset to climax offset (RQ #1)?

We found that gesture position (*G_position_rel*) was associated with *G_expressivity* [$\chi^2(3) = 20.902$, $p < 0.001$, with gestures that occurred later in a story showing higher *G_expressivity* values (see Table 3). Checking the random slope coefficients for each speaker revealed that 77% of speakers showed this positive association. Conditional pseudo-R^2 for this model was 0.036.

3.2 Does gesture expressivity predict specific EDA responses in story participants (RQ #2)?

We found that the probability of specific EDA responses was positively associated with *G_expressivity* [$\chi^2(1) = 18.046$, $p < 0.001$]. See Table 4 for an overview of model coefficients. Only random intercepts were included in this model due to singular fit when including random slopes. Note that while there is a large spread of intercepts across individual participants in Figure 3, indicating a large amount of variance across participants, nearly all of the fit lines show the same positive association. Conditional pseudo-R^2 for this model was 0.464.

3.3 How important is the contribution of gesture expressivity to emotional resonance compared to the contribution of other predictors of resonance (RQ #3)?

The Random Forest (*ntree* = 1,500, *mtry* = 3) that was constructed for emotional resonance (*EDA_G_resonance*) as outcome variable and the seven GEI parameters as well as six more variables as predictors (*G_quote*, *Sentiment*, *Protagonist*, *Group_compose*, *Group_size,* and *Recency*) exhibited a very good fit: according to a one-tailed exact binomial test, the model was significantly better than chance/baseline ($p < 0.001$), the (traditional) R^2 was 0.876, and McFadden's R^2 scored an excellent 0.386.

All predictors were found to impact *EDA_G_resonance*. Analysis of variable importance showed *Group_composition* to be by far the most impactful predictor, followed by *Sentiment*, *ND* (nucleus duration), *Recency, FO* (gesture force), *SZ* (gesture size), *G_quote* (gesture is co-quote), *HO* (gesture includes hold phase),

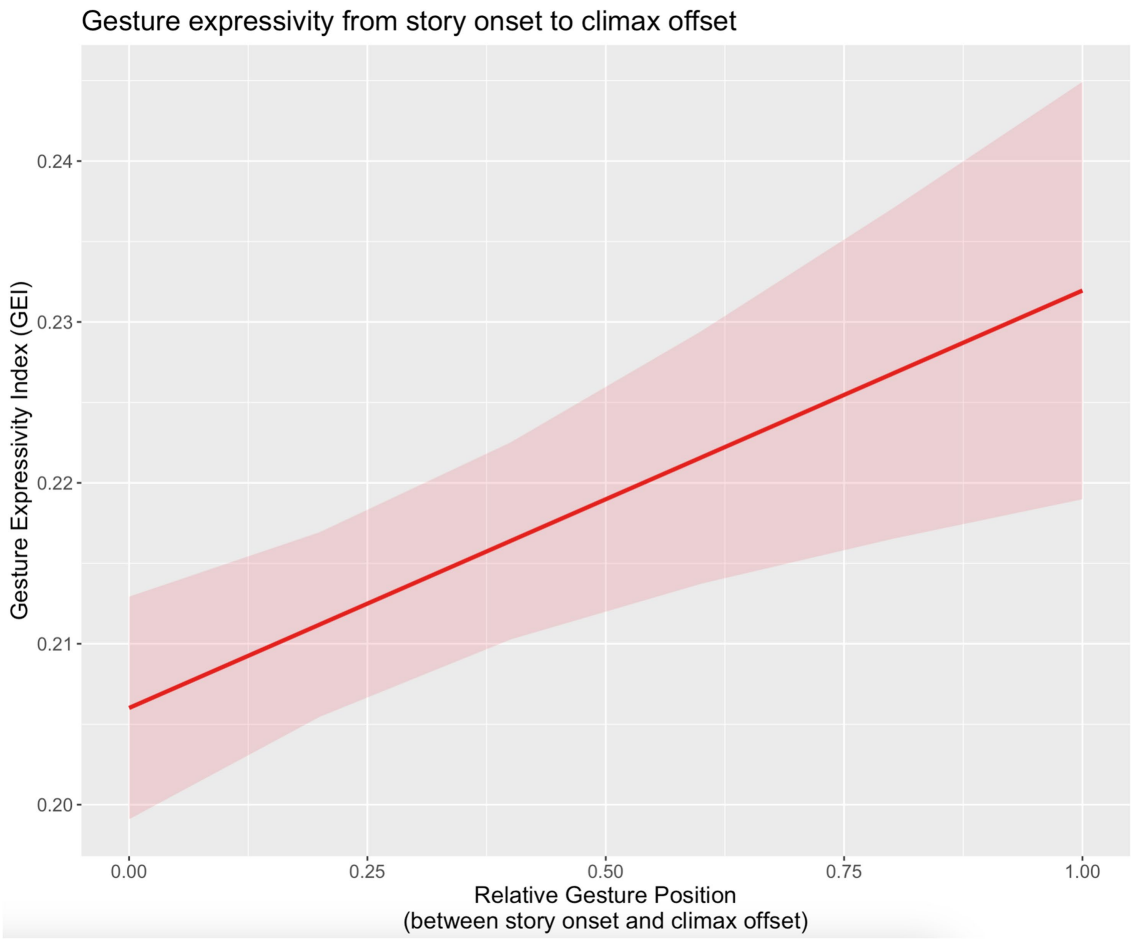


FIGURE 3
Linear mixed-effects regression model for gesture expressivity (*G_expressivity*) in storytellings (from story onset to climax offset); *G_position_rel*: relative positions of gestures in storytellings (values between 0 and 1).

Group_size, CV (gesture is character viewpoint), *SL* (silence during gesture), *Protagonist*, and *MA* (multiple articulators). While all variables had positive importance scores (indicating they contribute positively to model accuracy) the scores for *SL* (0.0013) and *MA* (0.0009) are very small (likely a reflection of their rarity; cf. Section 2.4.3) (See Figure 5).

Inspection of ICE plots strongly indicated combined effects of individual GEI parameters and other factors, including *Group_size* [the probabilities that gesture force (FO), size (SZ) and, respectively, nucleus duration (ND) impacts *EDA_G_resonance* were higher in triads] and *Group_compose* (the probabilities that these parameters impact *EDA_G_resonance* were much higher for all-men groups than for all-female and mixed groups). Figure 6 shows an ICE plot depicting the effect on emotional resonance of gesture size (*SZ*) interacting with group size (*Group_size*), while Figure 7 depicts the effect on emotional resonance of nucleus duration (*ND*) interacting with group composition (*Group_compose*). As can be seen from the plots, the effects of the two gesture kinematics (size and, respectively, nucleus duration) are much stronger for triads than for dyads on the one hand and for all-men groups than for all-female or mixed groups on the other. Note that these combined effects for group size and, respectively, group composition and gesture kinematics on emotional coupling were observed *consistently* across all seven gesture kinematics (FO, ND, SZ, HO, CV, MA, and SL).

4 Discussion

Overall, this study suggests that gesture expressivity increases over the course of a story and contributes to emotional resonance in conversational storytelling interaction.

Examination of RQ #1 demonstrated that story narrators' gestures become more expressive from story onset to climax offset. As the relation of gesture expressivity with conversational storytelling progression[4] has, to the best of our knowledge, not yet been examined

4 The notion of storytelling progression is tied to structural models of storytelling such as Labov's (1972) model and models underlying conversation-analytic analyses of storytelling. Their common denominator is the assumption that storytelling is a structured activity: "a story is not, in principle, a block of talk" (Jefferson, 1978, p. 245) but falls into "larger structures of talk" (Goodwin, 1984, p. 241) These "larger structures" are the story's "sections" (Labov, 1972),

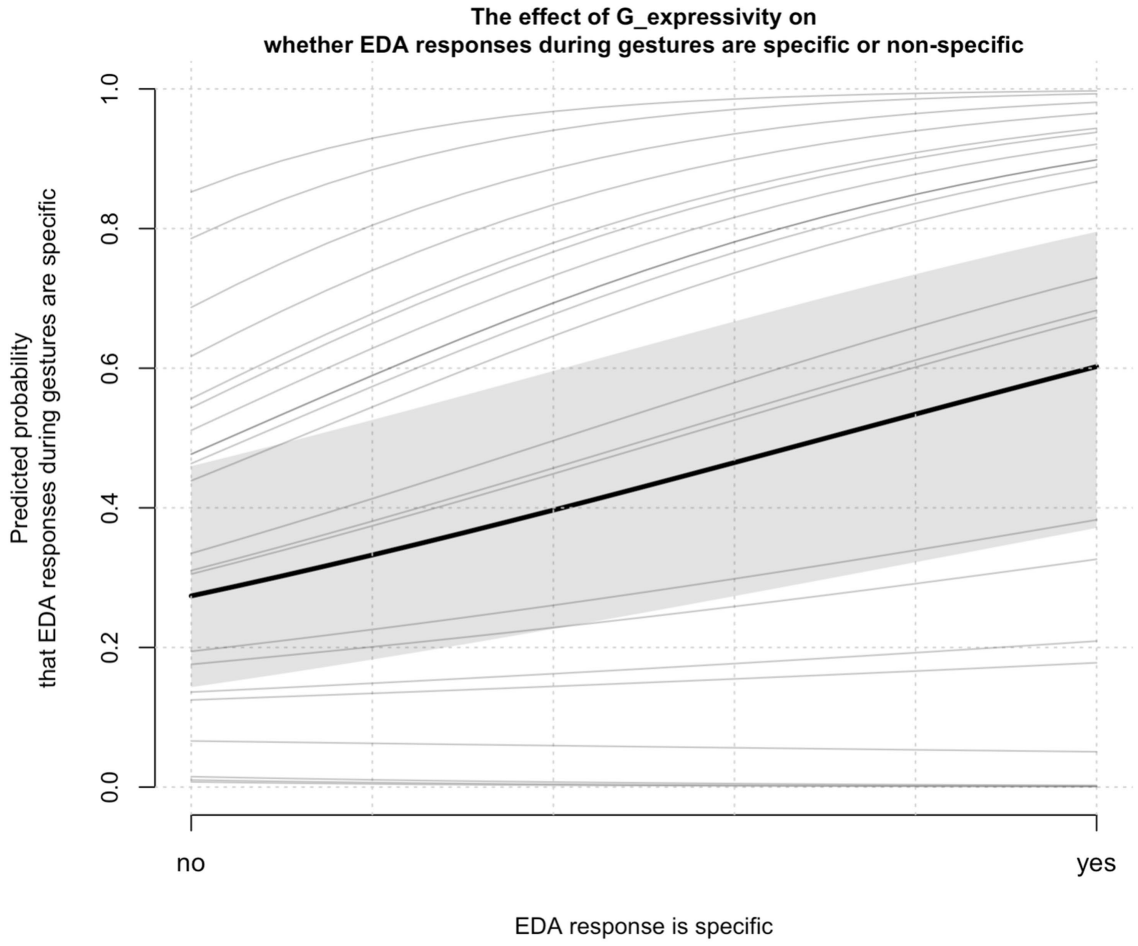


FIGURE 4
The effect of gesture expressivity (*G_expressivity*) on whether EDA responses are specific or not (*EDA_G_specific_responses_binary*); thin gray lines indicate individual participants.

statistically, this finding is the first of its kind. Gesture expressivity thus joins the group of expressive multimodal means, such as constructed dialog (e.g., Mayes, 1990), gaze alternation (Rühlemann et al., 2019) as well as intensity and pitch (Goodwin, 1984), that story narrators ratchet up to advance-project the imminent arrival at the story climax. Our finding adds further weight to the *Multimodal Crescendo Hypothesis* (Rühlemann, 2022), which posits that the story narrator's multimodal effort is *synchronized* with the storytelling's progression toward the Climax: "multimodal resources are deployed climactically such that they peak when the telling peaks reaching its key event, thus illuminating it brightly so that the event will be recognized as the key event at which displays of emotional contagion are relevant" (Rühlemann, 2022, p. 22). While this multimodal progression was present in most (77%) of storytellings, it should be noted that not all participants showed this same progression. This means that while there is evidence for a multimodal crescendo, not all speakers or stories will show this effect. Whether such differences are more speaker-specific or story-specific is an interesting avenue for future research, as this may be informative for how different types of stories utilize multimodal resources in different ways in order to build up to their climax, or whether speakers do not fully utilize visual signals to deliver their story. In the latter case, this would be a potentially fruitful avenue for investigating and/or training effective storytelling practices.

Addressing RQ #2, we found that increased gesture expressivity increases the probability of specific EDA responses (indicating emotion arousal) in the partcipants to the storytelling interaction. This finding is strong evidence that gestures in conversational storytelling can have an effect on the way the telling is *experienced* emotionally by the participants. Story narrators exploit the expressive claviature of gestures, skillfully varying and intensifying gesture kinematics to effectively change the way the recipients feel: far from merely listening to and comprehending what happened they get pulled into the story narrator's emotional orbit, potentially allowing them to partake in their emotions. This is consistent with the idea that multimodal expression enhances perceived emotions (Kelly and Ngo Tran, 2023), as well as the large body of research suggesting that the display of emotion, as in gestures, stimulates reciprocal emotional response (cf. Keltner and Ekman, 2000 and

"segments" (Jefferson, 1978) or "components" (Goodwin, 1984) including Preface, Background, and Climax as well as Post-completion sequences.

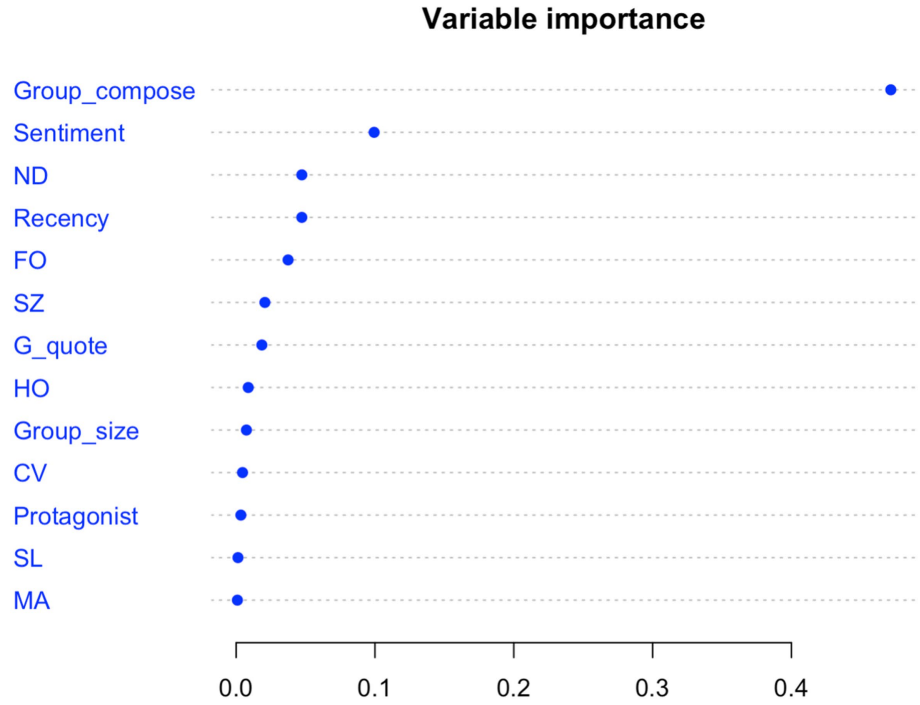


FIGURE 5
Overview of conditional variable importance in the Random Forest model predicting probability of specific EDA response. Individual predictor variables are given on the y-axis, while conditional variable importance is on the y-axis. Variable importance represents the mean decrease in accuracy of the model prediction when a given variable is removed.

references therein). It is also consistent with the observed tendency for people in interaction to continuously and non-consciously monitor and mimic the other's emotional expression and to "synchronize [their] expressions, vocalizations, postures, and movements and, consequently, to converge emotionally" (Hatfield et al., 1994, p. 5; Doherty, 1997, p. 149). The emotional contagion we observe in conversational storytelling interaction is clear evidence that storytelling has far deeper functions than just updating others so they know what happened; in storytelling, gestures (and other expressive means) can effectively evoke emotional arousal in their recipients, allowing for a more holistic experience of the story. Given that empathy comprises both cognitive empathy, "the intellectual/imaginative apprehension of another's mental state" (Lawrence et al., 2004, p. 911), and affective empathy playing out on the psycho-physiological level of emotion, the power of storytelling derives from the fact that it activates both dimensions of empathy. Affective empathy can be of two kinds, parallel and reactive. If the observer's empathic response matches that of the observed (your joy becomes my joy), the empathy is parallel; if the observer's response is complimentary (your distress becomes my compassion), the empathy is reactive. But, as noted, based on EDA only, we cannot distinguish kinds of empathy. Whether gestures evoke these *parallel or complimentary* emotional states in the recipients as conveyed or intended by the story narrator will require further research utilizing different methods.

Note, however, that specific EDA responses only indicate the arousal of emotion as such. They do not allow us to identify the kind of emotion the person is experiencing. To identify particular emotions, alternative psychophysiological metrics are required. For example, decelerating heart beat is indicative of sympathetic observers of other's sadness/distress, while the heartbeat of an observer with a self-focused personal distress reaction accelerates (Eisenberg et al., 1989 and references therein; cf. also Keltner and Ekman, 2000; Ekman et al., 1983).

Focusing on RQ #3, the analysis also demonstrated that gesture expressivity and the gesture dynamics that contribute to it allow conversational storytelling participants to resonate emotionally with one another by experiencing *simultaneous* emotion arousal. Considering that emotions are basic in the sense that they may have evolved "for their adaptive value in dealing with fundamental life tasks" (Ekman, 1999, p. 46) such as loss, danger, achievement, or fulfillment, being moved emotionally when the other is moved emotionally is, then, to resonate vis-à-vis any such fundamental life task. Darwin (1872) suggested that by sharing emotional states, individuals faced with such life tasks can bond, warn each other of danger, coordinate group activities, and, ultimately, enhance their chances of survival. In support of this claim, for example, Dunbar et al. (2016) demonstrated that, due to activation of the endorphine system, watching tragic films together increased not only social bonding but also tolerance of pain, thus making people effectively more resilient. Taking this evolutionary perspective, we can also speculate that emotional resonance in conversational storytelling would support the sharing of advice and (life) strategies, in the form of stories, where recognizing the emotions of events would be important. Our results suggest that gestures can support this resonance, and kinematic modulation of these gestures further contributes to the emotional resonance.

However, the analysis also demonstrated that gesture expressivity and the contributing gesture dynamics are just one set of factors in a

FIGURE 6
ICE plot of effect of interaction of Gesture size (*SZ*): Group size (*Group_size*) on Emotional resonance (*EDA_G_resonance*); values on the y-axis represent jittered means of predicted probabilities that emotional resonance is achieved (*EDA_G_resonance* = "yes").

complex web of factors that are *together* co-responsible for whether or not conversational storytelling participants get into synch emotionally with one another. Based on the analysis of variable importance, the analysis even suggested that gesture expressivity may not be the most impactful factor. Other non-gesture-related factors include, for example, the gender composition of the group of participants (by far the most important factor to emerge from the Random Forest model), the intensity of sentiment expressed in the co-gesture speech, whether the story events can be considered relevant given their recency and the co-presence of the protagonist, whether the story is told in dyads or triads, and whether the gesture is a co-quote gesture (i.e., whether it is part of the delivery of constructed dialog). Including the sentiment intensity of co-gesture speech, and seeing that gesture kinematics are still important in the model, further demonstrates that the synchronized affective responses of story narrator and listener cannot be explained by gestures simply appearing together with emotionally salient speech. Instead, the emotional kinematic modulation of gestures seems to play a role in affective alignment between individuals. So, whether storytelling in conversation fulfills its primary purpose—to facilitate a meeting of hearts—depends on a concert of factors.

Inspecting ICE plots, we also found evidence that in creating emotional coupling, gesture expressivity and its kinematic components enter into significant interactions with non-gesture-related situational factors. Such interactions include the interactions of group composition (all-female, all-male, and mixed) and, respectively, group size (dyads v. triads) on the one hand and all seven gesture kinematic parameters on the other. The former interactions suggest that the effects of the gesture kinematics on the achievement of emotional coupling are much greater in all-male groups than all-female or mixed groups. The latter interactions suggest that the influence of gesture dynamics on emotional synchrony are stronger in triadic than dyadic conversational storytellings. We will refrain here from commenting on the association between gender, gender composition, and emotional synchrony. The finding that group size—both in itself and in its interactions with gesture kinematics—impacts resonance suggests that the basic interactional organization of a conversation—whether it is between two or more people—matters fundamentally. Only very few studies have concerned themselves so far with the effects of group size. For example, Holler et al. (2021) found that response times in triads were shorter than dyads

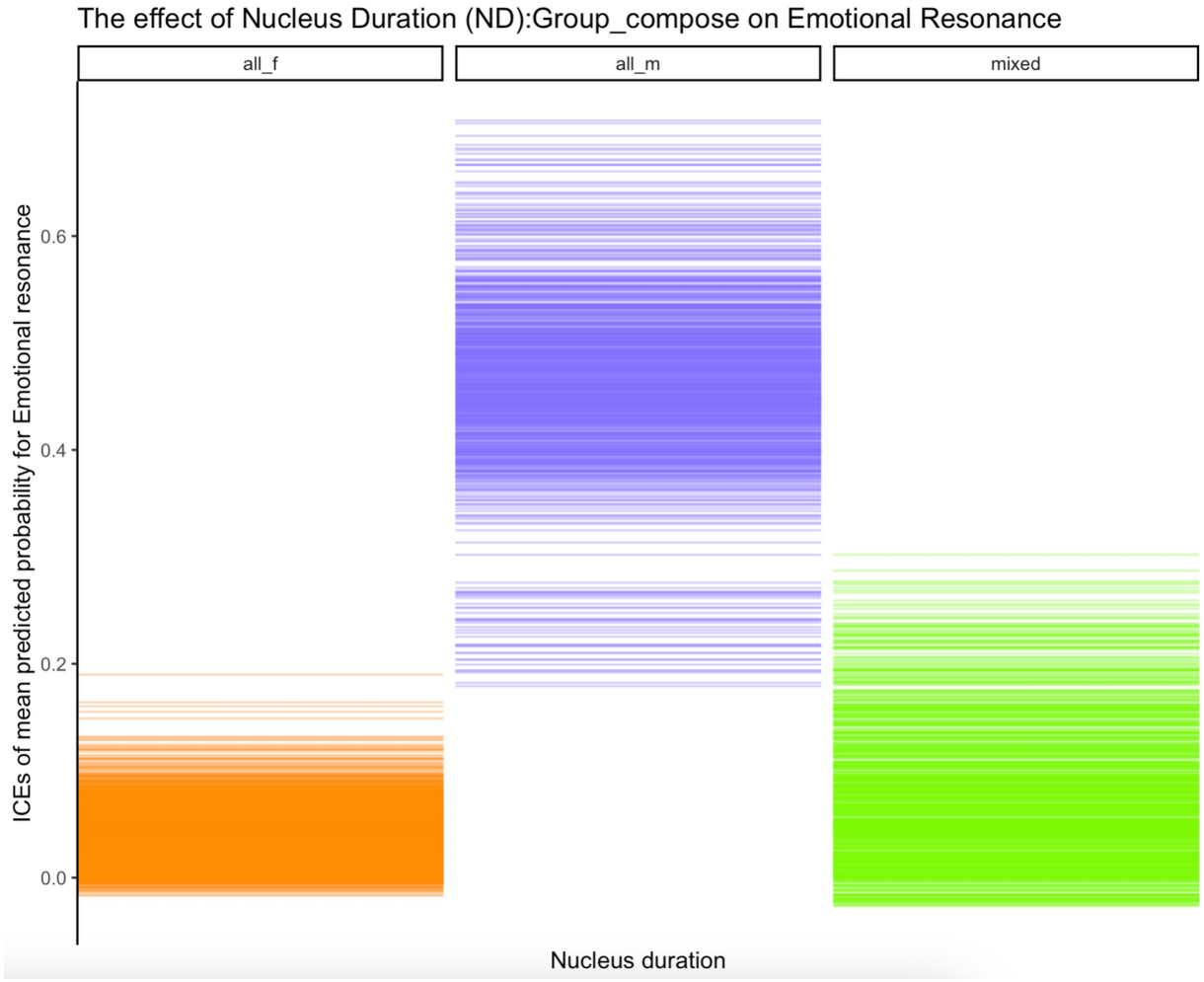


FIGURE 7
ICE plot of effect of interaction of Nucleus duration (*ND*):Group composition (*Group_compose*) on Emotional resonance (*EDA_G_resonance*); values on the y-axis represent jittered means of predicted probabilities that emotional resonance is achieved (*EDA_G_resonance* = "yes").

due to, the authors argue, competition. This study points to the possibility that group size affects interactional coupling dynamics far beyond just response timing.

When considering multimodal emotional expresssion and physiological synchrony, one important outcome of the Random Forest analysis is the inclusion of both gesture expressivity and speech sentiment as contributing to emotional arousal. Specifically, we show that while the intensity of sentiment conveyed in speech is associated with emotional arousal in the listener, the expressivity of co-occurring gestures also plays a role. This finding highlights the notion of *multimodal expressivity*, and is in line with previous findings that gestures can enhance the emotions conveyed by speech (Asalıoğlu and Göksun, 2023; Guilbert et al., 2021; Levy and Kelly, 2020). Finally, it is important to note that the present quantification of gesture expressivity is derived from a theoretically-motivated mix of kinematic and multimodal-embedding features. It may be useful to explore which other features contribute to an even more meaningful gesture expressivity index. For example, features worth examining in this resepect include fluidity and rhythmicity of movement (Hartmann et al., 2005; Pelachaud, 2009; Pouw et al., 2020; Trujillo et al., 2019).

5 Concluding remarks

Storytelling in conversation is an important "body-based way for instantiating a socioemotional connection with another" (Marsh et al., 2009, p. 334). This study has demonstrated that gestures play a key role in establishing that connection: story narrators use gestures skillfully varying their kinematic properties and expressive potentials. The effect of that kinematic virtuosity is emotional resonance: a momentary coupling of emotional affectedness in participants to storytelling—given the deep connections of emotions to life tasks, this coupling represents a powerful way of closing the inherent gap between bodies, brains, and hearts.

Specifically, this study produced three novel findings. First, the kinematic expressivity of gestures increases as the storytelling progresses toward the climax, following the general emotional

build-up conveyed through speech and following the basic climactotelic structure of storytelling. Second, increased gesture expressivity during storytelling increases the probability that participants to the storytelling experience specific EDA responses, that is, responses that reveal gesture-related emotional arousal. Third, gesture expressivity, in all its kinematic diversity, is one important factor contributing to the achievement of emotional resonance between storytelling participants; the most important factor, among a number of non-gesture-related linguistic and situational factors engendering emotional coupling, however, is the gender composition of the storytelling group. We also observed that all gesture kinematics substantially interact with group composition, and also group size, in engendering emotional resonance.

A limitation to this study is that despite the already large number of factors considered it may still not factor in all sources potentially influencing emotional resonance. Factors that future studies would necessarily have to take into account include the participants' interpersonal dynamics (whether they are strangers, friends, or romantic partners), paralinguistic prosody, which enables speakers to "achieve an infinite variety of emotional, attitudinal, and stylistic effects" (Wennerstrom, 2001, p. 200; cf. also Goodwin et al., 2012; for a study on paralinguistic synchrony see Paz et al., 2021) and also gaze: Kendon (1967), for example, observes hightened emotional arousal in phases of mutual gaze (cf. also Hietanen, 2018). Another factor to examine relates to the "Big Five" personality traits: for example, extroversion (v. introversion) has been shown to correlate with more sizable gestures (e.g., Mehl et al., 2007). Another, probably even more elusive, factor influencing emotional resonance is the extent to which people are *susceptible* to emotional resonance in the first place. The range of factors that may cause individual differences in susceptibility to emotional contagion include for example genetics, early experience, and personality characteristics (Doherty, 1997, p. 133).

Methodologically, this study opens up new avenues of multimodal corpus linguistic research by examining gesture at micro-analytic kinematic levels and using advanced machine-learning methods to deal with the inherent collinearity of multimodal variables. More good is expected to come from this fruitful combination of qualitative and quantitative research.

Data availability statement

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

Ethics statement

Ethical approval was not required for the studies involving humans because participants gave written informed consent stating their choices as to which of their data and in which form it could be used. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

CR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. JT: Formal analysis, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Deutsche Forschungsgemeinschaft grant number 497779797.

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

Asalıoğlu, E. N., and Göksun, T. (2023). The role of hand gestures in emotion communication: do type and size of gestures matter? *Psychol. Res.* 87, 1880–1898. doi: 10.1007/s00426-022-01774-9

Bailey, R. (2017). ""Electrodermal activity (EDA)" in The international Encyclopedia of communication research methods. eds. J. Matthes, C. S. Davis and R. F. Potter (Blackwell: Wiley), 1–15.

Bamberg, M. (2004). Talk, small stories, and adolescent identities. *Hum. Dev.* 47, 366–369. doi: 10.1159/000081039

Bartoń, K. (2009). Mu-MIn: Multi-model inference. R Package Version 0.12.2/r18. http://R-Forge.R-project.org/projects/mumin/

Beattie, G. (2016). Rethinking body language: How hand movements reveal hidden thoughts. London: Routledge.

Blackwell, N. L., Perlman, M., and Fox Tree, J. E. (2015). Quotation as multi- modal construction. *J. Pragmat.* 81, 1–17. doi: 10.1016/j.pragma.2015.03.004

Bradley, M. M., and Lang, P. J. (2007). "Emotion and motivation" in Handbook of psychophysiology. eds. J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (Cambridge: Cambridge University Press), 581–607.

Brone, G., and Zima, E. (2014). Towards a dialogic construction grammar: ad hoc routines and res- onance activation. *Cognit. Ling.* 25, 457–495. doi: 10.1515/cog-2014-0027

Chafai, N. E., Pelachaud, C., Pelé, D., and Breton, G. (2006). "Gesture expressivity modulations in an ECA application," in I*ntelligent Virtual Agents: 6th International Conference, IVA 2006, Marina Del Rey, CA, USA, August 21–23, 2006* (Berlin: Springer Berlin Heidelberg), 181–192.

Clark, H. H., and Gerrig, R. J. (1990). Quotations as demonstrations. *Language* 66, 764–805. doi: 10.2307/414729

Clift, R., and Holt, E. (2007). "Introduction" in Reporting talk. Reported speech in interaction (Cambridge: Cambridge University Press), 1–15.

Couper-Kuhlen, E. (2012). "Exploring affiliation in the reception of conversational complaint stories" in Emotion in interaction. eds. A. Peräkylä and M. L. Sorjonen (New York, NY: Oxford University Press), 113–146.

Culpeper, J. (2011). Impoliteness: Using language to cause offence. Cambridge: Cambridge University Press.

Dael, N., Goudbeek, M., and Scherer, K. R. (2013). Perceived gesture dynamics in nonverbal expression of emotion. *Perception* 42, 642–657. doi: 10.1068/p7364

Darwin, C. (1872). The expression of emotions in man and animals. London: John Murray.

Dawson, M. E., Schell, A. M., and Fillion, D. F. (2000). "The electrodermal system" in Handbook of physiology. eds. J. T. Ciacoppo, L. G. Tassinary and G. G. Berntson (New York, NY: Cambridge University Press), 200–223.

Doherty, R. W. (1997). Emotional contagion scale: a measure of individual differences. *J. Nonverbal Behav.* 21, 131–154. doi: 10.1023/A:1024956003661

Dunbar, R. I. M., Teasdale, B., Thompson, J., Budelmann, F., Duncan, S., van Emde, B. E., et al. (2016). Emotional arousal when watching drama increases pain threshold and social bonding. *R. Soc. Open Sci.* 3:160288. doi: 10.1098/rsos.160288

Duran, N. D., and Fusaroli, R. (2017). Conversing with a devil's advocate: interpersonal coordination in deception and disagreement. *PLoS One* 12:e0178140. doi: 10.1371/journal.pone.0178140

Ebisch, S. J., Aureli, T., Bafunno, D., Cardone, D., Romani, G. L., and Merla, A. (2012). Mother and child in synchrony: thermal facial imprints of autonomic contagion. *Biol. Psychol.* 89, 123–129. doi: 10.1016/j.biopsycho.2011.09.018

Eisenberg, N., Fabes, R. A., Miller, P. A., Fultz, J., Shell, R., Mathy, R. M., et al. (1989). Relation of sympathy and personal distress to prosocial behavior: a multimethod study. *J. Pers. Soc. Psychol.* 57, 55–66. doi: 10.1037/0022-3514.57.1.55

Eisenberg, N., and Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychol. Bull.* 94, 100–131. doi: 10.1037/0033-2909.94.1.100

Ekman, P. (1999). "Basic emotions" in Handbook of cognition and emotion. eds. T. Dalgleish and M. J. Power (New York, NY: John Wiley & Sons Ltd), 45–60.

Ekman, P., and Friesen, W. V. (1974). Detecting Deception from the Body or Face. Journal of Personality and Social Psychology. , 29, 288–298.

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system distinguishes between emotions. *Science* 221, 1208–1210. doi: 10.1126/science.6612338

Fusaroli, R., and Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive science*, 40, 145–171. doi: 10.1111/cogs.12251

Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Top. Cogn. Sci.* 1, 292–304. doi: 10.1111/j.1756-8765.2009.01020.x

Gates, K. M., Gatzke-Kopp, L. M., Sandsten, M., and Blandon, A. Y. (2015). Estimating time-varying RSA to examine psycho- physiological linkage of marital dyads. *Psychophysiology* 52, 1059–1065. doi: 10.1111/psyp.12428

Georgakopoulou, A. (1997). Narrative performances: A study of modern Greek storytelling. Amsterdam: John Benjamins.

Giles, H., Coupland, J., and Coupland, N. (Eds.) (1991). Contexts of accommodation. Cambridge: Cambridge University Press.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat.* 24, 44–65. doi: 10.1080/10618600.2014.907095

Goodwin, C. (1984). "Notes on story structure and the organization of participation" in Structures of social action: Studies in conversation analysis. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 225–246.

Goodwin, C. (1986). Between and within alternative sequential treatments of continuers and assessments. *Hum. Stud.* 9, 205–217. doi: 10.1007/BF00148127

Goodwin, M. H., Cekaite, A., and Goodwin, C. (2012). "Emotion as stance" in Emotion in interaction. eds. A. Peräkylä and M.-L. Sorjonen (New York, NY: Oxford University Press), 16–41.

Gries, S. T. (2021). Statistics for linguistics with R. *3rd* Edn. Berlin: Mouton de Gruyter.

Guilbert, D., Sweller, N., and Van Bergen, P. (2021). Emotion and gesture effects on narrative recall in young children and adults. *Appl. Cogn. Psychol.* 35, 873–889. doi: 10.1002/acp.3815

Gullberg, M., and Holmquist, K. (2002). "Visual attention towars gestures in face-to-face interaction vs. on screen" in Gesture and sign Langugae in human-computer interaction. eds. I. Wachsmuth and T. Sowa (Berlin/heidelberg: Springer), 206–214.

Halliday, M. A. K., and Matthiessen, M. I. M. (2004). An introduction to functional Grammar. *3rd* Edn. London: Edward Arnold.

Hartmann, B., Mancini, M., and Pelachaud, C. (2005). "Implementing expressive gesture synthesis for embodied conversational agents" in International gesture workshop (Berlin, Heidelberg: Springer), 188–199.

Hatfield, E., Cacioppo, J., and Rapson, R. (1994). Emotional Contagion. New York, NY: Cambridge University Press.

Hietanen, J. K. (2018). Affective eye contact: an integrative review. *Front. Psychol.* 9:1587. doi: 10.3389/fpsyg.2018.01587

Holler, J., Alday, P. M., Decuyper, C., Geiger, M., Kendrick, K. H., and Meyer, A. S. (2021). Competition reduces response times in multiparty conversation. *Front. Psychol.* 12:693124. doi: 10.3389/fpsyg.2021.693124

Holler, J., and Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends Cogn. Sci.* 23, 639–652. doi: 10.1016/j.tics.2019.05.006

Holley, J. W., and Guilford, J. P. (1964). A note on the G index of agreement. *Educ. Psychol. Meas.* 24, 749–753. doi: 10.1177/001316446402400402

Holt, E. (2007). "I'm eying your chop up mind': reporting and enacting" in Reporting talk. Reported speech in interaction. eds. E. Holt and R. Clift (Cambridge, MA: Cambridge University Press), 47–80.

Hsu, H.-C., Brône, G., and Feyaerts, K. (2021). When gesture "takes over": speech-embedded nonverbal depictions in multimodal interaction. *Front. Psychol.* 11:552533. doi: 10.3389/fpsyg.2020.552533

Hutto, C., and Gilbert, E. (2014). Vader: a parsimonious rule-based model for sentiment analysis of social media text. *Proc. Int. AAAI Conf. Web Soc. Media* 8, 216–225. doi: 10.1609/icwsm.v8i1.14550

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., et al. (2016). Cohesion and joint speech: right hemisphere contributions to synchronized vocal production. *J. Neurosci.* 36, 4669–4680. doi: 10.1523/JNEUROSCI.4075-15.2016

Jefferson, G. (1978). "Sequential aspects of storytelling in conversation" in Studies in the Organization of Conversational Interaction. ed. J. Schenkein (New York, NY: Academic Press), 219–248.

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction" in Conversation analysis. Studies from the first generation. ed. G. H. Lerner (Amsterdam: John Benjamins), 13–31.

Kelly, S. D., and Ngo Tran, Q. A. (2023). Exploring the emotional functions of co-speech hand gesture in language and communication. *Topics Cognit. Sci.* doi: 10.1111/tops.12657

Keltner, D., and Ekman, P. (2000). "Facial expression of emotion" in Handbook of emotions. eds. M. Lewis and J. Haviland-Jones. *2nd* ed (New York, NY: Guilford Publications, Inc), 236–249.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychol.* 26, 22–63. doi: 10.1016/0001-6918(67)90005-4

Kendon, A. (2004). Gesture: Visible action as utterance. Cambridge, MA: Cambridge University Press.

Kendon, A. (2017). Pragmatic functions of gestures: some observations on the history of their study and their nature. *Gesture* 16, 157–175. doi: 10.1075/gest.16.2.01ken

Kipp, M., Neff, M., and Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form. *Lang. Resour. Eval.* 41, 325–339. doi: 10.1007/s10579-007-9053-5

Koban, L., Ramamoorthy, A., and Konvalinka, I. (2019). Why do we fall into sync with others? Interpersonal synchronization and the Brain's optimization principle. *Soc. Neurosci.* 14, 1–9. doi: 10.1080/17470919.2017.1400463

Konvalinka, I., Sebanz, N., and Knoblich, G. (2023). The role of reciprocity in dynamic interpersonal coordination of physiological rhythms. *Cognition* 230:105307. doi: 10.1016/j.cognition.2022.105307

Kendon, A. (1973). The role of visible behavior in the organization of social interaction. in *Social Communication and Movement: Studies of Interaction and Expression in Man and Chimpanzee*, eds. M. Cranach and I. Vine (New York, NY: Academic Press), 29–74.

Koul, A., Ahmar, D., Iannetti, G. D., and Novembre, G. (2023). Interpersonal synchronization of spontaneously generated body movements. *IScience* 26:106104. doi: 10.1016/j.isci.2023.106104

Labov, W. (1972). Language in the Inner City. Oxford: Basil Blackwell.

Labov, W., and Waletzky, J. (1967). "Narrative analysis: Oral versions of personal experience" in Essays on the verbal and visual arts. ed. H. June (Seattle: University of Washington Press), 12–44.

Lawrence, E. J., Shaw, P., Baker, D., Baron- Cohen, S., and David, A. S. (2004). Measuring empathy: reliability and validity of the empathy quotient. *Psychol. Med.* 34, 911–920. doi: 10.1017/S0033291703001624

Levinson, S. C., and Holler, J. (2014). The origin of human multi-modal communication. *Philos. Trans. R. Soc. B* 369:20130302. doi: 10.1098/rstb.2013.0302

Levy, R. S., and Kelly, S. D. (2020). Emotion matters: the effect of hand gesture on emotionally valenced sentences. *Gesture* 19, 41–71. doi: 10.1075/gest.19029.lev

Li, C. L. (1986). "Direct and indirect speech: a functional study" in Direct and indirect speech. ed. F. Coulmas (Berlin: Mouton de Gruyter), 29–45.

Longacre, R. E. (1983). The grammar of discourse. New York, NY: Plenum Press.

Lykken, D. T., and Venables, P. H. (1971). Direct measurements of skin conductance: a proposal for standardization. *Psychophysiology* 8, 656–672. doi: 10.1111/j.1469-8986.1971.tb00501.x

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

Mathis, T., and Yule, G. (1994). Zero quotatives. *Discourse Process.* 18, 63–76. doi: 10.1080/01638539409544884

Mayes, P. (1990). Quotation in spoken English. *Stud. Lang.* 14, 325–363. doi: 10.1075/sl.14.2.04may

McFarland, D. H., Fortin, A. J., and Polka, L. (2020). Physiological measures of mother–infant interactional synchrony. *Dev. Psychobiol.* 62, 50–61. doi: 10.1002/dev.21913

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., and Pennebaker, J. W. (2007). Are women really more talkative than men? *Science* 317:82. doi: 10.1126/science.1139940

Mogan, R., Fischer, R., and Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J. Exp. Soc. Psychol.* 72, 13–20. doi: 10.1016/j.jesp.2017.03.009

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago: University of Chicago Press.

Norrick, N. R. (2000). Conversational narrative storytelling in everyday talk. Amsterdam: John Benjamins.

Ochs, E., and Capps, L. (2001). Living narrative. Cambridge, MA: Harvard University Press.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philos. Royal Soc. B Biol. Sci.* 369:20130296. doi: 10.1098/rstb.2013.0296

Özyürek, A. (2017). "Function and processing of gesture in the context of language" in Why gesture (Amsterdam: John Benjamins), 39–58.

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., et al. (2016). Interpersonal autonomic physiology: a systematic review of the literature. *Personal. Soc. Psychol. Rev.* 22, 1–43. doi: 10.1177/1088868316628405

Paz, A., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schectman, E., Gannot, S., Laufer-Goldshtein, B., et al. (2021). Intrapersonal and interpersonal vocal affect dynamics during psychotherapy. *J. Consult. Clin. Psychol.* 89, 227–239. doi: 10.1037/ccp0000623

Peeters, D., Chu, M., Holler, J., Hagoort, P., and Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* 27, 2352–2368. doi: 10.1162/jocn_a_00865

Pelachaud, C. (2009). Studies on gesture expressivity for a virtual agent. *Speech Comm.* 51, 630–639. doi: 10.1016/j.specom.2008.04.009

Peräkylä, A., Henttonen, P., Voutilainen, L., Kahri, M., Stevanovic, M., Sams, M., et al. (2015). Sharing the emotional load: recipient affiliation calms down the storyteller. *Soc. Psychol. Q.* 78, 301–323. doi: 10.1177/0190272515611054

Pouw, W., Jaramillo, J. J., Ozyurek, A., and Dixon, J. A. (2020). Quasi-rhythmic features of hand gestures show unique modulations within languages: evidence from bilingual speakers. *Listening* 3:61.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002

Romero, V., and Paxton, A. (2023). Stage 2: visual information and communication context as modulators of interpersonal coordination in face-to-face and videoconference-based interactions. *Acta Psychol.* 239:103992. doi: 10.1016/j.actpsy.2023.103992

Rühlemann, C. (2013). Narrative in English conversation: A corpus analysis of storytelling. Cambridge, MA: Cambridge University Press.

Rühlemann, C. (2019). Corpus linguistics for pragmatics (series 'Corpus Guides'). Abingdon: Routledge.

Rühlemann, C. (2022). How is emotional resonance achieved in storytellings of sadness/distress? *Front. Psychol.* 13:952119. doi: 10.3389/fpsyg.2022.952119

Rühlemann, C., and Ptak, A. (2023). Reaching below the tip of the iceberg: A guide to the Freiburg multimodal interaction corpus (FreMIC). *Open Ling.* 9:245. doi: 10.1515/opli-2022-0245

Rühlemann, C., Gee, M., and Ptak, A. (2019). Alternating gaze in multi-party storytelling. *J. Pragmat.* 149, 91–113. doi: 10.1016/j.pragma.2019.06.001

Rasenberg, M., Özyürek, A., and Dingemanse, M.. 2020. Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science* 44. doi: 10.1111/cogs.12911

Sacks, H. (1978). "Some technical considerations of a dirty joke" in Studies in the organization of conversational interaction. ed. J. Schenkein (New York, NY: Academic Press), 249–269.

Schegloff, E. A. (1997). Narrative analysis' thirty years later. *J. Narrat. Inq. Life Hist.* 7, 97–106. doi: 10.1075/jnlh.7.11nar

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inf.* 44, 695–729. doi: 10.1177/0539018405058216

Schmidt, R. C., and Richardson, M. J. (2008). "Dynamics of interpersonal coordination" in Coordination: Neural, behavioral and social dynamics (Berlin: Springer Berlin Heidelberg), 281–308.

Selting, M. (2010). Affectivity in conversational storytelling: an analysis of displays of anger or indignation in complaint stories. *Pragmatics* 20, 229–277. doi: 10.1075/prag.20.2.06sel

Selting, M. (2012). Complaint stories and subsequent complaint stories with affect displays. *J. Pragmat.* 44, 387–415. doi: 10.1016/j.pragma.2012.01.005

Siddle, D. (1991). Orienting, habituation, and resource allocation: an associative analysis. *Psychophysiology* 28, 245–259. doi: 10.1111/j.1469-8986.1991.tb02190.x

Silveira, P. S. P., and Siqueira, J. O. (2023). Better to be in agreement than in bad company: a critical analysis of many kappa-like tests assessing one-million 2x2 contingency tables. *Behav Res Methods*. 55, 3326–3347. doi: 10.3758/s13428-022-01950-0

Soulaimani, D. (2018). Talk, voice and gestures in reported speech: toward an integrated approach. *Discourse Stud.* 20, 361–376. doi: 10.1177/1461445618754419

Stec, K., Huiskes, M., and Redeker, G. (2016). Multimodal quotation: role shift practices. *J. Pragmat.* 104, 1–17. doi: 10.1016/j.pragma.2016.07.008

Stenström, A.-B., Andersen, G., and Hasund, I. K. (2002). Trends in Teenage Talk. Amsterdam: John Benjamins.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stivers, T/, and Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society*, 35, 367–392.

Tagliamonte, S., and Baayen, R. H. (2012). Models, forests and trees of York English: was/were variation as a case study for statistical practice. *Lang. Var. Chang.* 24, 135–178. doi: 10.1017/S0954394512000129

Tannen, D. (1986). "Introducing constructed dialog in Greek and American conversational and literary narrative" in Direct and indirect speech. ed. F. Coulmas (Berlin: Mouton de Gruyter), 311–332.

Tantucci, V., and Wang, A. (2021). Resonance as an applied predictor of cross-cultural interaction: constructional priming in mandarin and American English interaction. *Appl. Linguis.* 43, 115–146. doi: 10.1093/applin/amab012

Trujillo, J. P., and Holler, J. (2023). Interactionally embedded gestalt principles of multimodal human communication. *Perspect. Psychol. Sci.* 18, 1136–1159. doi: 10.1177/17456916221141422

Trujillo, J. P., Simanova, I., Bekkering, H., and Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: a Kinect study. *Cognition* 180, 38–51. doi: 10.1016/j.cognition.2018.04.003

Trujillo, J. P., Simanova, I., Bekkering, H., and Özyürek, A. (2020). The communicative advantage: how kinematic signaling supports semantic comprehension. *Psychol. Res.* 84, 1897–1911. doi: 10.1007/s00426-019-01198-y

Trujillo, J. P., Vaitonyte, J., Simanova, I., and Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: a toolkit for gesture and movement research. *Behav. Res. Methods* 51, 769–777. doi: 10.3758/s13428-018-1086-8

Tschacher, W., Greenwood, S., Weining, C., Wald-Fuhrmann, M., Ramakrishnan, C., Seibert, C., et al. (2024). Physiological audience synchrony in classical concerts linked with listeners' experiences and attitudes. *Sci. Rep.* 14:16412. doi: 10.1038/s41598-024-67455-2

van Ulzen, N. R., Lamoth, C. J. C., Daffertshofer, A., Semin, G. R., and Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neurosci. Lett.* 432, 88–93. doi: 10.1016/j.neulet.2007.11.070

Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. Royal Soc. B Biol. Sci.* 369:20130292. doi: 10.1098/rstb.2013.0292

Wallbott, H. G. (1998). Bodily expression of emotion. *Eur. J. Soc. Psychol.* 28, 879–896. doi: 10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

Wennerstrom, A. (2001). The music of everyday speech: Prosody and discourse analysis. Oxford: Oxford University Press.

Wharton, T., Bonard, C., Dukes, D., Sander, D., and Oswald, S. (2021). Emotion and relevance. *J. Pragmat.* 181, 259–269. doi: 10.1016/j.pragma.2021.06.001

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). *Elan: a professional framework for multimodality research. in Proceedings of LREC 2006.* (Genoa: LREC).

Xu, S., and Lorber, M. F. (2014). Interrater agreement statistics with skewed data: evaluation of alternatives to Cohen's kappa. *J. Consult. Clin. Psychol.* 82, 1219–1227. doi: 10.1037/a0037489

frontiers | Frontiers in Psychology

TYPE Brief Research Report
PUBLISHED 07 January 2025
DOI 10.3389/fpsyg.2024.1499192

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Ushba Rasool,
Zhengzhou University, China
Listyani Listyani,
Satya Wacana Christian University, Indonesia
Al Ryanne Gatcho,
Hunan Institute of Science and Technology, China

*CORRESPONDENCE
Leslie Huishan Li
✉ leslie.hs.li@connect.um.edu.mo

RECEIVED 20 September 2024
ACCEPTED 18 December 2024
PUBLISHED 07 January 2025

CITATION
Li LH (2025) Intersubjectivity and co-constructed framings: students' role-play talks in online English-speaking sessions. *Front. Psychol.* 15:1499192. doi: 10.3389/fpsyg.2024.1499192

COPYRIGHT
© 2025 Li. 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.

Intersubjectivity and co-constructed framings: students' role-play talks in online English-speaking sessions

Leslie Huishan Li*

Department of English, Faculty of Arts and Humanities, University of Macau, Macau SAR, China

This study investigates how lower-level English language learners achieve and maintain intersubjectivity and navigate through co-constructed framings during context-embedded tasks such as improvised role-play. In language education settings, activities associated with interactions often reflect multi-layered participant orientations beyond mere linguistic execution. From the perspectives of ethnomethodology and conversation analysis, linguistic actions are effective only when they are intersubjectively understood: the complexity of an activity necessitates corresponding levels of intersubjectivity for smooth progression. A close examination of sequence development in role-play activities shows how intersubjectivity is preserved as interlocutors engage in aligning and affiliative moves to avoid potential disruptions and maximize mutual contributions. Interlocutors' progressive inputs are integrated as intersubjectively sustained elements of ongoing interaction framings only when collaboratively developed by the participating parties. The analysis also addresses the impact of online communication tools, acknowledging their increasingly essential role in online teaching. The results suggest the need for a dynamic concept of 'framing,' replacing 'frame,' and recommend that conversation analysis should account for the multi-layered contexts of surrounding activities.

KEYWORDS

classroom interaction, online learning, student interaction analysis, role-play activity, conversation analysis

1 Introduction and literature review

1.1 CA, English language teaching, and student interaction

Conversation Analysis (CA), with its strength in detailing the nuances of social interaction, has been extensively applied to English language teaching, exploring both the "What" and "How" of the field [see Waring (2019), for an overview]. In defining "What to be taught" in English language teaching (ELT), CA-informed research has significantly contributed to the (re)conceptualization of interactional competence as the primary educational goal [examples include Barraja-Rohan (2011), Hall (2018), and Harumi (2023)]. It has also critically evaluated textbook design, ensuring alignment with real-life communication practices (e.g., Fredagsvik, 2023; Gardner, 2000; Wong, 2007). Additionally, CA research bridges theoretical advancements with practical applications in teaching by closely examining classroom instruction. CA-informed classroom discourse studies have scrutinized teachers' strategies for managing participation. CA-informed classroom discourse studies have examined teachers' strategies for managing participation (e.g., Fagan, 2012; Hosoda and Aline, 2013) and delivering instructions (e.g., Gosen et al., 2024; Markee, 2015; Seedhouse, 2008), thus informing "how teaching is conducted."

Moreover, CA works have demonstrated changes in portraying classroom interactions from teacher-fronted question-answer-comment sets to a nexus of interrelated speech exchange systems (Markee, 2000; Tai, 2023). These changes have been promoted by research on students' execution of interactional tasks in the classroom or where "learning-in-and-as-interaction" (Koschmann, 2013, p. 2) occurs. Such tasks are often designed in ways that afford the target language for communicative purposes and prompt learners to produce intended interactions (Huth, 2011). CA researchers have then investigated turn-to-turn behaviors and envisioned learning as local and contingent occasions. Previous CA studies discussed how learners collaboratively navigate diverse task trajectories (e.g., Hellermann and Doehler, 2010) and how different task stages as observable social processes become loci of learning (e.g., Markee and Kunitz, 2013; Doehler and Eskildsen, 2022).

1.2 CA, role-play, and applications in educational settings

Conversation analysts were among the earliest to investigate role-play interactions (Francis, 1989; Sharrock and Watson, 1985). These studies revealed that role-play is structured by participants' culturally bound reasoning protocols (Sharrock and Watson, 1985). Recent CA studies recognized that elicited role-play interactions share interactional features as in real-life conversations (Sikveland et al., 2023; Stokoe, 2013). In educational settings, role-play data can be used to examine how social actions are accomplished through learners' talk-in-interaction. This allows for conducting assessments and analyzing the needs of language learners (Youn, 2020). In ELT settings, role-play is a widely applied task type that requires highly context-dependent performance.

Okada (2010) studied candidates' performance in role-play tasks during oral language proficiency tests and found that participants display competencies in talk-in-interaction mechanisms (e.g., turn-taking and sequence organization) that closely resemble those found in actual conversations. Informed by task-based pragmatic needs analysis (Youn, 2018), Youn (2020) investigated participants' organizations of proposal sequences in role-plays designed for speaking assessment. While approving language teachers' application of role-plays as pedagogical tasks, Youn (2020) also reported relatively abrupt opening turns by lower-level learners, recommending pre-teaching necessary linguistic resources for organizing actions.

1.3 Overview and the present study

Previous research has acknowledged the advantages of engaging language learners in classroom activities in which the target language is designed for specific communicative purposes (Dos Santos, 2020; Lee, 2000). Moreover, CA research proved that interactants in role-play activities discursively accomplish social actions and deploy talk-in-interaction mechanisms that are highly similar to those in real-life conversations (see 1.2 above). Despite bearing such potential, existing studies show scarce attention to adopting role-play in language and communication development. Though a refined framework, the Conversation Analytic Role-play Method (CARM) proposed by Stokoe (2014) focused specifically on professional training, such as for language professionals and healthcare providers (Church and Bateman, 2020; Niemants et al., 2023). On learner interactions in ELT, the few existing CA studies analyzed role-play interactions only in oral tests or assessment tasks from the lenses of interactional competence and assessment criteria (Havadar and Balaman, 2024; Youn, 2020).

The present study aims to fill such gaps by examining learners' role-play conversations in non-assessment-related class activities, focusing on how interactants collaboratively maximize mutual construction and advance the fluent progress of the conversations. While Youn's (2020) study suggested that role-plays might not be appropriate for lower-level learners due to the requirement for managing contextual performances, findings from this study demonstrate how lower-level English language learners utilize communication resources and successfully maintain intersubjectivity. Data presented in this article are generated from an online English-speaking course designed for lower-level adult learners and are analyzed using conversation analysis. The course was conducted during a pandemic quarantine period through an online meeting platform; hence, the influence of online communication was also examined.

1.4 Data collection

The original dataset was generated from a series of online English language courses for lower-level adult learners. The participants consist of lower-intermediate level [A1-A2 CEFR, see British Council (2024)] Chinese learners who participated in weekly grammar and speaking sessions to improve their oral proficiency. In the speaking sessions, the learners engaged in group or pair tasks tailored to elicit peer conversations, thereby practicing the knowledge acquired in corresponding grammar sessions. Most of the tasks encompass topics closely related to the learners' daily lives and prompt them to simulate their routine social interactions in the target language. The online courses were conducted through an e-platform developed by the course provider. It should be noted that during the role-play activity, the students chose not to turn on their cameras, and only the performing pair would unmute themselves; the instructor also stayed camera-off and muted during each pair's play process to allow undistracted interaction. Therefore, the target segments generated audio-only pair-talk data. The audio recordings were considered to be of a high standard by colleagues during multiple data sessions. For audio recordings in the dataset, written consent from relevant participants was acquired, all data excerpts were fully de-identified through thorough pseudonymization, and the presented excerpts went through participant member-checking for anonymity.

Data analyzed in this study are drawn from a set of 10-week speaking sessions attended predominantly by junior undergraduate students. The excerpts presented in this article are drawn from three learner pairs' co-constructed talks in an improvised role-play activity (see Table 1), in which they were asked to role-play a "police officer interrogating potential suspect" scene (see Figure 1 below)[1]. The instructor led the learners through some basic inquire-and-answer expressions taught in the previous grammar session, whereas this activity had no pre-set expressions or plots and aimed to elicit learners'

1 The animated illustration on the instruction page (Figure 1) is acquired from the open-access and license-free website Freepik.com. Freepik.com

TABLE 1 Basic information of the selected student pairs.

(Pseudonym initials)	Role division	Main plots	Duration
Pair 1: A and B	A as police, B as a suspect	Alibi debate; witness reference	5 min 40 s
Pair 2: P and J	P as police, J as a suspect	Scene affirming; lawyer reference	3 min 20 s
Pair 3: Z and N	Z as police, N as a suspect	Framing another suspect	3 min 35 s

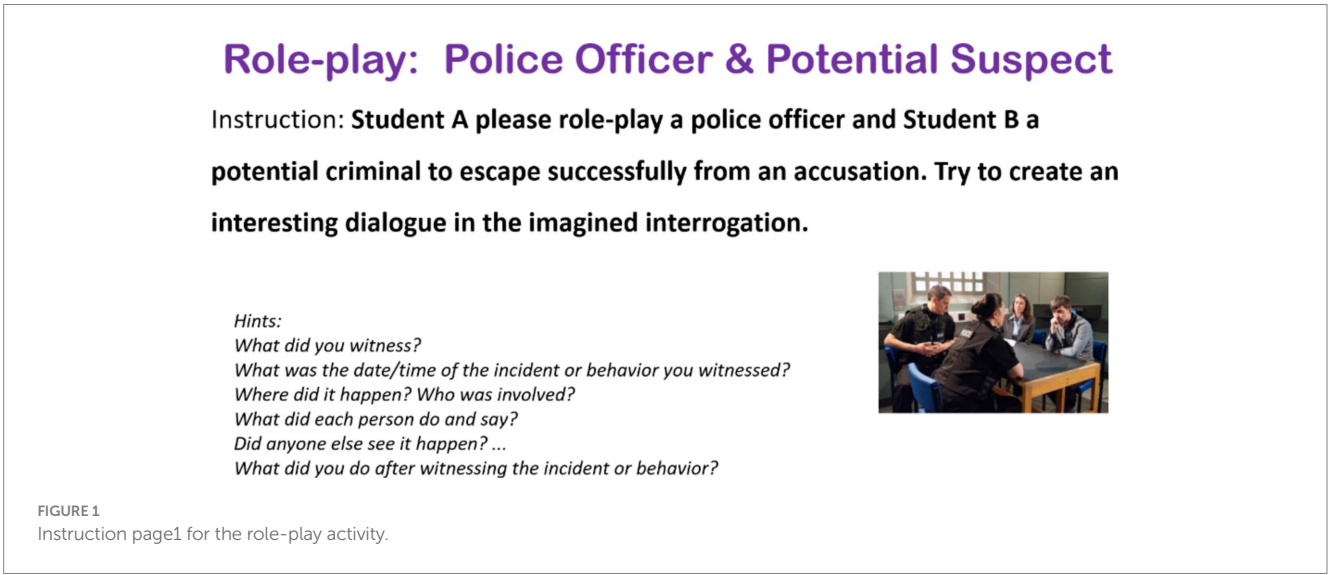


FIGURE 1
Instruction page1 for the role-play activity.

naturally occurring and unscripted interaction. The learners were given 15 min for discussion and rehearsal, during which the instructor circulated within the classroom to provide assistance when required. Each paired role-play lasted 3–6 min, and the selected learner pairs showed comparatively balanced mutual contribution during the plays.

1.5 Data analysis

From the CA perspective, linguistic actions have effects only when intersubjectively understood as so – the more complex an activity becomes, the more intricate intersubjectivity is required for it to proceed (Enfield and Sidnell, 2021). Highly contextual classroom activities such as role-plays could be notably challenging for lower-level learners as they are required to use a work-in-progress language. Previous CA studies on role-play activities in ELT primarily focused on the assessment of learners' interactional competence and with deficit assumptions on lower-level learners' performance in organizing such conversations. In contrast, data analysis in this study looks at how lower-level learners maintain mutual intersubjectivity by deploying contextual resources.

Data analysis also reveals the need for and adopts the concept of "framing" to replace "frame" (MacLachlan and Reid, 1994). In these role-play conversations, interlocutors assume dual roles as characters in the role-play and task performers in the language class, engaging with coexisting realms of contextualization, or "frames": the conversation in which the play occurs, the play activity itself as an in-situ performance, and the played event unfolding in imaginary space and time (Young, 1987). Such coexisting realms are dynamic and contingently constructed by interlocutors' moment-by-moment collaborative interactions, and there could naturally be several competing frames in the same talk. To analyze such dynamic, ever-changing, and multi-layered contexts and activities in conversations, "framing" is adopted instead of the more or less static notion of frame (Linell and Thunqvist, 2003).

The dataset was transcribed verbatim following Jefferson's (2004) CA transcription conventions (see Supplementary Appendix A). The transcriptions underwent three rounds of detailed revisions and two rounds of peer review in data sessions with CA researchers. After finalizing the transcription, I conducted a data exploration with unmotivated-looking, maximizing emergent interactional patterns from the data while avoiding deliberately picking out excerpts to fit into pre-determined concepts or frameworks (ten Have, 2007). I first analyzed randomly selected data excerpts to acquire a general picture of interactional mechanism patterns, such as sequence organization, turn-taking, repair, and laughter. This process identified recognizable sequential groups in terms of how intersubjectivity and conversation advancement are collaboratively maintained under online and audio-only conditions. Grasping the general picture of the interactional mechanisms and recognizable sequences allowed me to identify and organize the interlocutors' representative patterns across events, suggesting the need for a more dynamic concept of framing. Then, a collection of examples of similar phenomena was developed through several rounds of inductive explorations throughout the dataset. Excerpts presented below are further selected based on intelligibility quality and representativeness in elucidating the analysis focuses above.

2 Results

As illustrated above, in this role-play activity, students were asked to work in pairs, improvise, and produce an "interrogation" role-play conversation. In the following analysis, the presented excerpts are selected from three student pairs (see Table 1), with one pair (A and B) of particular focus due to data richness and representativeness.

2.1 Improvising role-play task online: achieve and maintain intersubjectivity

Excerpt 1 A & B

```
7  A: °O[kay]
8  B:   [Hey] cops. why take me here,
9  A: .hh (.) You ^don't know why?
10   .hh okay. WHAT are you doing in that day.
11   (1.4)
12 B: .h I'm: (0.3)((tongue clicking)) happy:-
13   ((tongue clicking)) I'm having my happy song in (.) KTV
14   with my school friend.
```

The AB pair's role-play sequences begin with B's first turn in line 8, which explicitly draws upon the task prompt (the police-suspect interrogation) and proffers a relevant topic ("Why take me here?"). In the first turn construction unit (henceforth TCU), B's stress of "Hey" signals an initiation of the upcoming sequences, while "cops" contextually cues A as both the assigned role (the police) and the next speaker. The second TCU then completes the topic with a question, inviting A's answer to expand the role-play opening upon B's completed turn. A's response picks up B's cue and orients to the projected continuation. A's rhetorical question (line 9) displays a receipt of both the police officer role and the story expansion expectation while prefacing a stance that B, as the suspect in the role-play story, should provide the alibi statement. The question in line 10 then elicits B's initiating alibi sequences, which B picks up and plays, following the projected continuation direction again.

As the role-play story unfolds turn by turn, the paired interlocutors adopt multiple strategies to prevent potential conversational problems and sustain the collaborative progression of the role-play sequences. The major strategies include longer pauses, absolute priority of self-repair, and analeptic tying. The lengthy pauses, instead of signaling non-alignment, delay responses to maximize collaborative contribution to the advancement of the talks. For example, in line 11 of Excerpt 1 above, the longer, uninterrupted pause allows B to construct prior-turn-based sequences, with A tacitly cooperating. The interlocutors also confirm the completion of previous turns through longer pauses, considering the absence of nonverbal communicational resources and the possibility of network delay (e.g., line 45 of Excerpt 2/line 12 of Excerpt 3 below). Correspondingly, the struggling self-repairs (line 6 and line 10 of Excerpt 3) are not intervened because the repairing results constitute essential, understandable, and expandable points for the next turn. Further, interlocutors use repetition as an analeptic tying device to invoke earlier-produced elements as resources for achieving interactional work (Mlynář, 2020). As Sacks (1992) noted, the first part of a tying point cannot be identified until the second part is formulated. Here, "KTV" and "sleep more" (lines 39 and 46 of Excerpt 2) are the tying points evoked for both double confirming a transitional element in the previous turn as well as strategically delaying while projecting the upcoming response based on that element. The local past is thus becoming the intersubjectively sustained constituent of ongoing interaction. Taken together, these ongoing sequences are co-constructed through the interlocutors' close monitoring of each other's turn construction, displaying nuanced and in-situ mutual understandings (Enfield and Sidnell, 2021). Because only speech acts can be observed during these online talks, interlocutors purposely adopt pause, repair, and tying strategies to maximize mutual construction. Intersubjectivity is thus achieved and maintained through purposive aligning and affiliative moves as both in-play characters and peer task performers (Stivers, 2008).

Excerpt 2 A & B

```
37 A: Okay: (0.4) BEFORE the KTV where are you go,
38   (1.0)
39 B: .hh (0.8) Before the ^KTV¿ stay at ^home (0.8)
40    in the morning. (0.6)
41    [I hav]ing (.) too much drink. (.) Um=
42 A: [hmm]
43 B: = (( tongue clicking)) at the day before the night.
(0.5)
44    So I've- have to sleep more.
45   (1.4)
46 A: Sleep more. .hh
47 B: °Yeah
48 A: =Just you:- just what- you go owh? (0.5)
```

Excerpt 3 P & J

```
5 Z: So: (.) Do you- do you ^know that
6     your ^neighbors, Mr.- (.) Mr.- uh: Mr.£^Smith£
7     be killed. Last night?
8   (0.4)
9 N: Sorry I don't know. I just saw e#gh#: woman
10    who closed. (.) .hh Ugh:: Who-who wear the
11    wright clothes on the streets.
12   (1.0)
13 Z: So: you only- (.) see the ^ladys.
14   (0.6)
15 N: Ye[ah.]
```

2.2 Improvised role-play talks: co-constructed and competing framings

Upon maintaining intersubjectivity, the interlocutors also frame coexisting contexts and perform alongside these ongoing frames-in-frames or competing framings (Linell and Thunqvist, 2003). In Excerpt 1 above, B's initial turn recognizes A's epistemic status as the police, thus enacting activity roles and signaling a shift into the specific interaction framing – an interrogation role-play. A's response, on the one hand, carries on the simulated interrogation framing with prosodic features as the police (lines 9 and 10), whereas on the other hand, it implicitly indicates the playful task framing as A reciprocates the informal appellation "cops" (as opposed to the formal form "sir").

In Excerpts 4 and 5 above, interlocutors playing the police are to invalidate the suspect's alibi by introducing and improvising a third

party. A's self-repair from "I" to "we" (line 57 of Excerpt 4) and Z's "witness" invocation (line 22, Excerpt 5) contribute to the "interrogation" framing by formulating the institutional stance (as "the police"). Moreover, features of troubles "you- uh you guys go- uh > go" (lines 57–60 of Excerpt 4) and repairs "But I hear- I have heard"(lines 21–24 of Excerpt 5) also indicate the ongoing improvised dialog framing where actors have to constantly produce frame-consistent contingencies, as well as the broader learning interaction where learners struggle at language production.

Returning to what happened prior to line 37 in the AB pair's conversation, on the one hand, the turns are primarily structured as question-and-answer adjacency pairs. On the other hand, the incremental expansions and displayed stances (lines 28, 32, 34 of Excerpt 6 below) based on prior turns indicate how the interlocutors frame the talk with stored realities beyond the ongoing task and the classroom. Such is also how scenes are kept moving in improvised dialogues, in which an actor proposes a new development to the play frame. That proposal becomes a constituent of the ongoing framing only when evaluated and embellished by the other actor. In other words, the actors' turn-by-turn, co-constructed interactional framings enable and constrain their next actions (Sawyer, 2003).

Excerpt 6 A & B

```
26   Okay. Then, which one KTV you go.
27   (1.8)
28 B: UV UV ktv. A[t half past thruu.
29 A:              [UV KTV.
30 B: Yeah.
31   (0.6)
32 A: But there: no camera can see you why,
33   (1.5)
34 B: .hh ((tongue clicking))Because I am having a situation
35    =I have to having my number two. So I go to my- go to
36    the toilet (.) in publi. Yeah.
37 A: Okay: (0.4) BEFORE the KTV where are you go,
38   (1.0)
39 B: .hh (0.8) Before the ^KTV¿ stay at ^home (0.8)
40    in the morning. (0.6)
```

As Goffman (1974) noted, discourse frames depend on participants attending to interactions with stored information from previous experiences. Interaction-advancing orientations are thus both turn-generated as the conversation unfolds and dependent on pre-existing expectations brought by the interactants. While the interlocutors collaboratively co-construct emergent framings through mutually interpretable contextual resources, their expression retrievals are also compatible with the broader framing of their conversation-for-learning and classroom interactions. Conversation sequences are thus simultaneously orienting to and compatible with multiple layers of competing framings. Rather than simply shifting between frames, participants blend or even "embed one within another" through such interdiscursivity (Gordon, 2008, p. 323).

2.3 "Moving between" framings: aligning laughter and playfulness

Cases exist, however, in which improvisational players suddenly move out of the ongoing frame and turn to certain metapragmatic actions – what Goffman (1974) referred to as frame-breaking moves. In these instances, discursively marked and abrupt transitions arise, and the players often explicitly deviate from their assigned roles within the primary improvisational framing.

Of Goffman's particular interest are those frame-breaking moves that are emotionally charged. In Excerpt 7 below, B, the suspect, defends his alibi by delegitimizing the witness's epistemic status and thus invalidating the statement against him. As discussed earlier, the witness is also an improvised adding element by A, the student, and enacted by A, the police, while it is closely consistent with and expectable following the focal "interrogation" framing. B's "idiot" utterance (line 62) not only initiates his alibi defense along the focal framing but also noticeably activates the playful task framing, which has also been concomitantly present. This turn is treated by A as a laughable point out of the primarily ongoing interrogation framing. A burst into laughter (line 63) not so much as A the police but more so as A the student or other roles/identities.

Excerpt 7 A & B

```
62 B: Why, (0.8) [Beca]use he is an idiot. (0.6) [He have
some
63 A:            [Yeah]                              [czi heh
.hhh
64 B: (0.4)He [h a v e] some um: (0.2)
65 A:         [°.h(h)h]
66 B: sick in his mind=He always will (.) uh forget anything
67   =So um maybe ^sometime (.) .h when he feel ^nervous
68   he will make up some stories=So don't believe him.
69   (.) Cops.
70 A: Ehm. You think (.) I will believe you?
71    (0.2)
72 B: .hhh (0.4)Yeah: (.) I can show you (0.5) He has uh-
73    seen the (0.3) doctor at the= (1.1)
74 A: eh-hhehh-h
75 B: =£l-last(h)week£. And the doctor said (.)
76    he had some problem in his mind and he always forget
77    something.
```

However, laughter requires alignment or appraisal between the participants involved to function properly (Glenn, 2003). In contrast, B only minimally caters to it through short pauses and a stressed voice (lines 64 and 66–69, Excerpt 7 above) before quickly returning to the interrogation framing. Hence, B's utterances are not specifically designed to make laughter relevant. Although A then attempts to refrain and abandon the laughter (lines 65 and 70) and focus again on the focal framing along B's commitment, his laughter abruptly reprises (line 74) while B finally joins the laughter (line 75). Therefore, the out-of-frame moves, or what Schegloff (2001) called nonseriousness, extend across multiple turns.

In contrast, N's out-of-frame laughter (line 39 of Excerpt 8) quickly receives Z's immediate alignment (line 40), this time with Z attempting to return to the police role and the main focal framing (line 42). However, Z soon joins N in laughter, realigning with the lighter, playful tone. Earlier, in Excerpt 7, A initiates laughter (line 63, Excerpt 7) in overlap with B's ongoing turn. Here, N, as the current speaker (line 39, Excerpt 8), laughs first, marking her turn with nonseriousness, which shifts its original sequential implications within the interrogation framing. In response, Z's

prosodically inflected "silence" (line 42, Excerpt 8), followed by a short pause, signals an orientation back to reprised playfulness.

Excerpt 8 Z & N

```
33 Z: But according to the <^cameras>,
34     We- we ^viewed- (0.3)
35     >The camera has-< have shown that you have-
36     (0.3) go:- go into the man's °house
37     (0.5) So be honest. What do you do there.
38    (0.6)
39 N: Uh I just- £on my way home£.
40 Z: Uh-huh. [Y o u]=
41 N:          [Really]
42 Z:= have the right to be ^silence,(.) thhehh
43 N: £Yeah£.
```

Hence, in these seemingly outer framings, participants tend to behave non-seriously, less consistently, and/or more playfully than in the ongoing focal framings. This is conveyed through contextualization cues such as laughter and aspiration (Gumperz, 1992). Laughter may explicitly frame prior or ongoing turns, and to some extent the upcoming pairs, toward playful ends, thus indicating potential framing boundaries and shifts in behavior. Rather than not carrying the sequential implications within the primary, more serious framing, these turns allow flexible responses as the relevant next along multiple framings (Holt, 2016). Therefore, as discussed in previous sections, participants are, in fact, simultaneously orienting to multiple layers of partially competing yet partially parallel framings.

3 Discussion and conclusion

This project investigates how paired English-language learners achieve and maintain intersubjectivity while moving in between co-constructed framings during online role-play talks. A close examination of sequence development made visible how intersubjectivity is meticulously maintained in improvised talks, even by relatively lower-level learners. Interlocutors are oriented toward aligning and affiliative moves to prevent potential problems, maximize mutual contribution, and form coherent interactions. Such moves are notably present as the interlocutors are engaged in online and audio-only communication. As they are only temporally co-present and not physically, the compromised voice delivery (e.g., volume, quality, and network latency) and the absence of non-verbal interpretable communicational resources (e.g., gaze and gestures) naturally alter the procedural infrastructure of interaction. Interactional features such as longer pauses, instant alignment, and inclinations to repair strategies show how collaborative meaning-making orients to the maximization of mutual intersubjectivity and the progressivity of conversations (Moorhouse et al., 2021).

Furthermore, the ongoing sequences in improvised role-play dialogues exhibit the interlocutors' mutual understanding, which is maintained and evolved moment by moment throughout the conversation. For example, when the responses correspond with previous turns and reflect either alignment or disagreement. In these improvised dialogues, the interlocutors' progressive contributions are integrated as intersubjectively sustained elements of ongoing interaction frameworks only when collaboratively co-developed by the other party involved. In other words, intersubjectivity serves both as a resource for and an outcome of the interaction (Heritage, 2012), existing as both "chronic" within the temporal scope of the current interaction and "diachronic," utilizing accumulated experiences and resources accumulated from various contexts (Enfield and Sidnell, 2021; Goodwin, 2018).

The availability of such resources depends on the interlocutors' ability to bring past experiences into conversations and how these experiences are collectively interpreted and utilized for meaning-making during the interaction. In this study, the interlocutors' use of communicative resources and strategies emerges from their collective knowledge of, among other factors, interrogation techniques, task-oriented playfulness, and the impact of online communication.

The limitations of this study should be noted. The analysis focuses on audio-only data. Though this may have more accurately explained the turn-by-turn pauses and stresses, the nonverbal, embodied communicational resources (gaze and gesture) normally contributing to social interactions are not included. The audio-only nature of the data also compromised the analysis of the influence across modalities brought by technology as the medium, such as the use of platform functions, the arrangement of online meeting layout, and participant interaction with the interface. In addition, though the analysis presents nuanced meaning-making dynamics, the findings do not comprehensively represent the range of sequence features for this activity type or learner levels due to the sample size. As Walsh (2011) argued, "If we want to look for evidence of learning, we should begin by focusing on the words and interactions of the learner" (p. 188). For language learners, formulating sequenced utterances in coherent and orderly manners is an important step toward advanced proficiency levels (Ortega and Byrnes, 2008).

To conclude, this study proposes CA approaches to examine the turn-by-turn performance of learner conversations and the nuanced forms of meaning-making. First, data analysis and the findings suggest the need for a dynamic concept of framing in replacement of frame to address the ever-changing and emergent nature of interlocutors' moment-by-moment, co-constructed meaning-making (MacLachlan and Reid, 1994), depicting multi-layered contexts of surrounding parallel activities (Linell and Thunqvist, 2003). Moreover, one important direction for current and future research is to investigate the constraints and affordances of the constantly advancing technological mediums of interpersonal interactions when dealing with naturally occurring conversation data.

Further, this study recommends context-embedded classroom activities such as role-play to help analyze learners' interactional competencies that are present in actual conversations. For language teachers in particular, observing and facilitating learners' conversations in context-embedded activities could allow understanding of how collaborative meaning-making and interpretation are achieved with the target language's varying and often early-stage proficiency. Through deploying a range of communicational resources, learners enact different social and participation roles in the interactions, creating multiple layers to the classroom participation framework and, more importantly, displaying their epistemic knowledge in class (Monfaredi, 2023). Instead of prescriptive and deficit perspectives on task appropriateness or assessing interactional competence for lower-level learners, future studies could take the perspective of resource deployment and negotiation in collaborative meaning-making across different learner levels.

Understanding the accessibility and understandability of resource deployment in meaning-making, or the actual outcome of conversations, would require investigating the differing past trajectories of the parties engaged. While CA approaches primarily focus on the turn-by-turn nuances in talks without making assumptions about broader social settings, studying interactions in specific institutional settings indeed brings attention to the influence brought by wider social and cultural contexts. Future research on language communication and development may consider combining CA with other bottom-up and meticulously descriptive approaches, taking emic perspectives to explore possibilities of mobilizing learner interlocutors' multifaceted built-in knowledge and resources. This could benefit from cultivating dispositions toward communication-by-repertoire instead of separate language systems and balancing participation dynamics in the classroom and other interactions. Moreover, the differences in personal trajectories and thus unequal access to resources, communicational or others, should be carefully treated in deference to inclusivity and diversity in education.

Data availability statement

The datasets presented in this article are not readily available because the full dataset may include identifiable data. Requests to access the datasets should be directed to the corresponding author leslie.hs.li@connect.um.edu.mo.

Ethics statement

Ethical approval was not required for the studies involving humans because ethical review and approval were not required for this study involving human participants in accordance with local legislation. Written informed consent was obtained from the individuals involved. All data presented in this article are pseudonymized and fully de-identified. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) to publish any potentially identifiable images or data in this article.

Author contributions

LL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Part of the findings have been presented at the Digital Meeting for Conversation Analysis 2022. I am grateful to Dr. Younhee Kim for her constructive feedback on an earlier version of the manuscript. My special thanks also go to the reviewers and editors for their efforts and highly valuable suggestions.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2024.1499192/full#supplementary-material

References

Barraja-Rohan, A.-M. (2011). Using conversation analysis in the second language classroom to teach interactional competence. *Lang. Teach. Res.* 15, 479–507. doi: 10.1177/1362168811412878

British Council. (2024). "Understanding English levels". Available at: https://learnenglish.britishcouncil.org/english-levels (Accessed June 03, 2024).

Church, A., and Bateman, A. (2020). Conversation analytic role-play method (CARM) for early childhood teacher education. *Teach. Dev.* 24, 652–668. doi: 10.1080/13664530.2020.1820371

Doehler, P. S., and Eskildsen, S. W. (2022). Emergent L2 grammars in and for social interaction: introduction to the special issue. *Mod. Lang. J.* 106, 3–22. doi: 10.1111/modl.12759

Dos Santos, L. M. (2020). The discussion of communicative language teaching approach in language classrooms. *J. Educ. E-Learn. Res.* 7, 104–109. doi: 10.20448/journal.509.2020.72.104.109

Enfield, N., and Sidnell, J. (2021). "Intersubjectivity is activity plus accountability" in Oxford handbook of human symbolic evolution. eds. N. Gontier, A. Lock and C. Sinha (Oxford: Oxford University Press).

Fagan, D. S. (2012). Dealing with unexpected learner contributions in whole group activities: an examination of novice language teacher discursive practices. *Classroom Discourse* 3, 107–128. doi: 10.1080/19463014.2012.716621

Francis, D. (1989). "Game identities and activities: some ethnomethodological observations" in Communication and simulation: From two fields to one theme. eds. D. Saunders and D. Crookall (London: Multilingual Matters).

Fredagsvik, M. S. (2023). The challenge of supporting creativity in problem-solving projects in science: a study of teachers' conversational practices with students. *Res. Sci. Technol. Educ.* 41, 289–305. doi: 10.1080/02635143.2021.1898359

Gardner, R. (2000). Resources for delicate manoeuvres: learning to disagree. *Austr. Review Appl. Linguist. Suppl. Ser.* 16, 31–47. doi: 10.1075/aralss.16.03gar

Glenn, P. (2003). Laughter in interaction (studies in interactional sociolinguistics). Cambridge, UK: Cambridge University Press.

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Cambridge, Massachusetts: Harvard University Press.

Goodwin, C. (2018). Co-operative action. Cambridge, UK: Cambridge University Press.

Gordon, C. (2008). A(p)parent play: blending frames and reframing in family talk. *Lang. Soc.* 37, 319–349. doi: 10.1017/S0047404508080536

Gosen, M. N., Willemsen, A., and Hiddink, F. (2024). Applying conversation analysis to classroom interactions: students' 'oh'-prefaced utterances and the interactional management of explanations. *Classroom Discourse* 1–19, 1–19. doi: 10.1080/19463014.2024.2397127

Gumperz, J. J. (1992). "Contextualization and understanding" in Rethinking context: Language as an interactive phenomenon. eds. A. Duranti and C. Goodwin (Cambridge: Cambridge University Press).

Hall, J. K. (2018). From L2 interactional competence to L2 interactional repertoires: reconceptualising the objects of L2 learning. *Classroom Discourse* 9, 25–39. doi: 10.1080/19463014.2018.1433050

Harumi, S. (2023). Classroom silence and learner-initiated repair: using conversation analysis–informed material design to develop interactional repertoires. *TESOL J.* 14:e704. doi: 10.1002/tesj.704

Havadar, E. E., and Balaman, U. (2024). Preliminaries in paired role play L2 interactional competence assessment tasks. *Classroom Discourse* 15, 143–160. doi: 10.1080/19463014.2023.2240898

Hellermann, J., and Doehler, P. S. (2010). On the contingent nature of language-learning tasks. *Classroom Discourse* 1, 25–45. doi: 10.1080/19463011003750657

Hepburn, A., and Bolden, G. (2017). "Transcribing aspiration and laughter" in Transcribing for social research. eds. A. Hepburn and G. Bolden (Thousand Oaks: SAGE Publications Ltd), 65–80.

Heritage, J. (2012). The epistemic engine: sequence organization and territories of knowledge. *Res. Lang. Soc. Interact.* 45, 30–52. doi: 10.1080/08351813.2012.646685

Holt, E. (2016). Laughter at last: playfulness and laughter in interaction. *J. Pragmat.* 100, 89–102. doi: 10.1016/j.pragma.2016.04.012

Hosoda, Y., and Aline, D. (2013). Two preferences in question–answer sequences in language classroom context. *Classroom Discourse* 4, 63–88. doi: 10.1080/19463014.2013.783497

Huth, T. (2011). Conversation analysis and language classroom discourse. *Lang. Linguist. Compass* 5, 297–309. doi: 10.1111/j.1749-818x.2011.00277.x

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction" in Conversation analysis: Studies from the first generation. ed. G. Lerner (Amsterdam, Netherlands: John Benjamins).

Koschmann, T. (2013). "Conversation analysis and learning in interaction" in In *the encyclopedia of applied linguistics*, edited by C (Chapelle: Blackwell).

Lee, J. F. (2000). Tasks and communicating in language classrooms. New York: McGraw-Hill.

Linell, P., and Thunqvist, D. P. (2003). Moving in and out of framings: activity contexts in talks with young unemployed people within a training project. *J. Pragmat.* 35, 409–434. doi: 10.1016/S0378-2166(02)00143-1

MacLachlan, G., and Reid, I. (1994). Framing and interpretation. Carlton: Melbourne University Press.

Markee, N. (2000). Conversation analysis. New York: Routledge.

Markee, N. (2015). "Giving and following pedagogical instructions in task-based instruction: an ethnomethodological perspective" in International perspectives on ELT classroom interaction. eds. C. J. Jenks and P. Seedhouse (London: Palgrave Macmillan).

Markee, N., and Kunitz, S. (2013). Doing planning and task performance in second language acquisition: an ethnomethodological respecification. *Lang. Learn.* 63, 629–664. doi: 10.1111/lang.2013.63.issue-4

Mlynář, J. (2020). "I'll tell you later on": Proleptic and analeptic tying devices in oral history interviews. *Narrat. Inq.* 30, 161–184. doi: 10.1075/ni.18020.mly

Monfaredi, E. (2023). Student-initiated storytelling in classroom interaction. *Classroom Discourse* 14, 411–439. doi: 10.1080/19463014.2022.2103007

Moorhouse, B. L., Li, Y., and Walsh, S. (2021). E-classroom interactional competencies: mediating and assisting language learning during synchronous online lessons. *RELC J.* 54, 114–128. doi: 10.1177/0033688220985274

Niemants, N., Hansen, J. P. B., and Stokoe, E. (2023). "The conversation analytic role-play method" in The Routledge handbook of public service interpreting. eds. L. Gavioli and C. Wadensjö (London: Routledge).

Okada, Y. (2010). Role-play in oral proficiency interviews: interactive footing and interactional competencies. *J. Pragmat.* 42, 1647–1668. doi: 10.1016/j.pragma.2009.11.002

Ortega, L., and Byrnes, H. (2008). "Theorizing advancedness, setting up the longitudinal research agenda" in The longitudinal study of advanced L2 capacities. eds. L. Ortega and H. Byrnes (New York: Routledge).

Sacks, H. (1992) in Lectures on conversation I–II. eds. H. Sacks and G. Jefferson (Oxford: Wiley-Blackwell).

Sawyer, R. S. (2003). Improvised dialogues: Emergence and creativity in conversation. Santa Barbara: Greenwood Publishing Group.

Schegloff, E. A. (2001). Getting serious: joke → serious 'no'. *J. Pragmat.* 33, 1947–1955. doi: 10.1016/S0378-2166(00)00073-4

Seedhouse, P. (2008). "Learning to talk the talk: conversation analysis as a tool for induction of trainee teachers" in Professional encounters in TESOL: Discourses of teachers in teaching. eds. S. Garton and K. Richards (Basingstoke: Palgrave Macmillan).

Sharrock, W. W., and Watson, D. R. (1985). "Reality construction" in second language-learning (L2) simulations. *System* 13, 195–206. doi: 10.1016/0346-251X(85)90033-8

Sikveland, R. O., Moser, T., Solem, M. S., and Skovholt, K. (2023). The effectiveness of the conversation analytic role-play method (CARM) on interactional awareness: a feasibility randomized controlled trial with student teachers. *Teach. Teach. Educ.* 129:104136. doi: 10.1016/j.tate.2023.104136

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stokoe, E. H. (2013). The (in) authenticity of simulated talk: comparing role-played and actual interaction and the implications for communication training. *Res. Lang. Soc. Interact.* 46, 165–185. doi: 10.1080/08351813.2013.780341

Stokoe, E. H. (2014). The conversation analytic role-play method (CARM): a method for training communication skills as an alternative to simulated role-play. *Res. Lang. Soc. Interact.* 47, 255–265. doi: 10.1080/08351813.2014.925663

Tai, K. W. H. (2023). Multimodal conversation analysis and interpretative phenomenological analysis: A methodological framework for researching Translanguaging in multilingual classrooms. *1st* Edn. London: Routledge.

ten Have, P. (2007). Doing conversation analysis: A practical guide (2nd ed.). London: SAGE.

Walsh, S. (2011). Exploring classroom discourse: Language in action. Abingdon, England: Routledge.

Waring, H. Z. (2019). "The what and how of English language teaching: conversation analysis perspectives" in Second handbook of English language teaching. ed. X. Gao (Cham: Springer International Handbooks of Education).

Wong, J. (2007). "Answering my call: a look at telephone closings" in Conversation analysis and language for specific purposes. eds. H. Bowles and P. Seedhouse (Bern: Peter Lang).

Youn, S. J. (2018). Task-based needs analysis of L2 pragmatics in an EAP context. *J. Engl. Acad. Purp.* 36, 86–98. doi: 10.1016/j.jeap.2018.10.005

Youn, S. J. (2020). Managing proposal sequences in role-play assessment: validity evidence of interactional competence across levels. *Lang. Test.* 37, 76–106. doi: 10.1177/0265532219860077

Young, K. (1987). Taleworlds and telling-worlds: The phenomenology of narrative. Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 15 January 2025
DOI 10.3389/fpsyg.2024.1463449

Check for updates

OPEN ACCESS

EDITED BY
Geert Brône,
KU Leuven, Belgium

REVIEWED BY
Omid Khatin-Zadeh,
University of Electronic Science and Technology of China, China
Roey Gafter,
Ben-Gurion University of the Negev, Israel

*CORRESPONDENCE
Anna Inbar
inbara9@gmail.com

RECEIVED 11 July 2024
ACCEPTED 30 November 2024
PUBLISHED 15 January 2025

CITATION
Inbar A (2025) Interactional practices accomplished by index-finger pointing directed at the addressee in Hebrew face-to-face interaction.
Front. Psychol. 15:1463449.
doi: 10.3389/fpsyg.2024.1463449

COPYRIGHT
© 2025 Inbar. 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.

Interactional practices accomplished by index-finger pointing directed at the addressee in Hebrew face-to-face interaction

Anna Inbar*

Department of Hebrew Language, Levinsky-Wingate Academic College, Tel Aviv, Israel, University of Haifa, Haifa, Israel

This study uses Hebrew data to examine the practices accomplished by index-finger pointing toward the addressee, with a focus on interactional purposes beyond merely indexing the reference. The data were taken from the *Haifa Multimodal Corpus of Spoken Hebrew*, which consists of video recordings of naturally occurring casual conversations collected between 2016 and 2023. By employing the methodologies of interactional linguistics and multimodal conversation analysis, the study elaborates on the social actions that are accomplished via this gesture, showing that pointing at the addressee in Hebrew talk-in-interaction can be explained from different perspectives. The study suggests that non-referential pointing primarily serves as an attention-drawing device. However, similar to other gestural or verbal attention-drawing devices, in some contexts, the gesture can also be considered to be a cue whereby conveying a negative stance or displaying epistemic authority is recognized. Additionally, it can be employed as an abrupt way of interrupting or as an attempt to elicit a response from the addressee.

KEYWORDS

pointing, attention-drawing device, multimodal stance-taking, negative stance, epistemic authority

1 Introduction

The prototypical pointing gesture (*cf. pure pointing*, Kendon, 1980; *pointing-out*, Lakoff, 1987; *canonical pointing*, Langacker, 2008) is a bodily movement toward a target in order to direct someone's attention to it (e.g., Eco, 1976, p. 119; Clark, 2003; Cooperrider, 2023; Cooperrider et al., 2018; Kendon, 2004; Kita, 2003). The target of pointing is presumed to be visual and present in the speech situation; if the pointing gesture is co-produced with speech, it is taken for granted that what is pointed to is in some sense identical to what is simultaneously referred to in speech (*cf.* Clark et al., 1983). An extended index finger is often considered to be the origin of pointing and is connected to the target via an imaginary line or trajectory (e.g., Enfield, 2009; Kita, 2003; McNeill et al., 1993). However, the preference for pointing with the index finger is not universal (e.g., Cooperrider et al., 2018; Wilkins, 2003), and pointing comprises a much broader range of bodily actions involving the thumb, hands, extended arm, head, face (e.g., lip-pointing), and objects (e.g., Cooperrider, 2023; Kendon, 2004). Moreover, even within cultures in which the gesture prototypically takes the form of an extended index finger, other forms can also be used, potentially revealing functional differences (*cf.* Kendon, 2004). The gesture is usually characterized by a "post-stroke hold" (Kita et al., 1998) or "stasis" (Cooperrider, Forthcoming), a brief visual suspension of the gesture from a

dynamic to a static position before being retracted or beginning the next gesture (*cf.* Kendon, 1980, 2004; Bressem and Ladewig, 2011).

Whereas prototypical pointing is used to make a reference to entities that are physically present in the immediate space of the interaction, such as an object, a person, a location, or a direction (e.g., Clark, 2003; Kendon, 2004; Kita, 2003), pointing can also be directed toward a seemingly empty space to provide new references (e.g., *abstract deixis*, McNeill, 1992; McNeill et al., 1993; *Deixis am Phantasma*, Stukenbrock, 2014). In addition, the pointed-at object can metonymically represent the intended (discourse) referent (e.g., Haviland, 2000; Levinson, 2006). As noted by Cooperrider (2014), such metonymic pointing is relatively well attested in the ethnographic literature on pointing, leading researchers to suggest that this type of metonymy "may be a pervasive feature of pointing in real-world settings" (p. 3).

While some studies of pointing gestures have addressed their deictic referential function (e.g., Clark, 2003; Kendon, 2004; Kita, 2003), other studies have revealed interactional practices that are accomplished by pointing (e.g., Bangerter, 2004; Cooperrider, 2011, 2014, 2016; Enfield et al., 2007; Goodwin, 2003; Healy, 2012; Hindmarsh and Heath, 2000; Holler, 2010; Mondada, 2007, 2012, 2014; Streeck, 2017; Yasui, 2023). For example, Mondada (2007) observed a work meeting involving a team of agronomists and computer scientists, and noted that the speakers used pointing gestures directed toward maps and other documents as a turn-taking device: In the turn-initial position, pointing indicated incipient speakership; in the pre-turn-initial position, pointing could be used as a claim for the next turn before the prior turn had been completed.

The trajectory of a pointing gesture can single out one of the participants in a conversation. As prototypical pointing is presumed to bring the recipient's attention to a pointed-at entity (e.g., Kita, 2003), the pointing at the addressee raises some questions, such as under what circumstances the speaker would request the addressee to pay attention to themselves, or whether the addressee is the true target of the pointing. Several studies have shown that a pointed-at co-participant can stand in an (apparent) metonymic relation of speaker for utterance (e.g., Ishino, 2009); for example, when the speaker cites what the conversational participant has just said (e.g., Bavelas et al., 1992). Other studies have shown that English-speakers may use a pointing gesture to indicate agreement with a pointed-at person (e.g., Healy, 2012). Some scholars (e.g., Bavelas et al., 1995; Enfield et al., 2007) have suggested that, in the course of interaction, the gesture can be used to specify the addressee of an utterance in order to elicit their response. However, pointing at the addressee appears to be a frequent phenomenon in dyadic face-to-face interactions (as the current study attests), in which singling out a co-participant as the intended recipient of a certain action is irrelevant.

The current study explores interactional practices, beyond indexing the reference, that are accomplished by index-finger pointing directed at the addressee by Hebrew speakers. The following sections first introduce the data and methodology (section 2). Subsequently, in order for the target phenomenon to emerge as a distinct one, several cases in which pointing at the addressee is used for deictic referential functions are provided (section 3). Non-referential pointing at the addressee—the cases in which the relationship between what is pointed at and what is said is not straightforward—is then elaborated on (section 4). Following this, the findings are discussed, showing that these pointing gestures appear to differ functionally from the canonical case (section 5). Finally, the study is summarized, and concluding remarks are provided (section 6).

2 Data and methodology

The data are drawn from the *Haifa Multimodal Corpus of Spoken Hebrew* (Maschler et al., 2024), which comprises approximately 22 h of video recordings of naturalistic conversations among friends and family members that were collected between 2016 and 2023. Informed consent for the collection and publication of the data was obtained from all of the participants. The participants were filmed in natural settings, including their homes, cafés, and workplaces, during casual conversations of approximately 30 min to 2 h. Following the setup of the camera and recording device, the researcher exited the environment, allowing the participants to interact freely without providing any instructions. This design facilitated uninstructed dialogue, enabling the participants to discuss the topics of their choice and fostering authentic social interaction. The present study is based on approximately nine and a half hours (571 min) of talk from 13 conversations—involving 30 speakers in total (nine dyadic, two triadic, and two quadratic conversations)—that were recorded between 2017 and 2021.

Tokens of addressee-directed index-finger pointing gestures were searched manually, excluding those used for indexing reference, such as when they were coordinated with utterances that included indexing the second person or when the pointed-at co-participant stood in an (apparent) metonymic relation of speaker for utterance or performer for action (see Section 3). The collection of addressee-directed index-finger pointing gestures, produced without any verbal reference to the addressee and not analyzed as metonymic pointing, comprised 81 tokens. These findings reveal that non-referential pointing directed at the addressee occurred at least every 5.3 min on average; ambiguous cases, such as those involving indexing the second person while also occurring in the contexts found to be associated with non-referential pointing, were not included in this count.[1]

To elaborate on the actions that were accomplished via such gestures, I employ the methodologies of interactional linguistics (Couper-Kuhlen and Selting, 2018) and multimodal conversation analysis (e.g., Goodwin, 2018; Mondada, 2016). The analysis will consider the position of the gesture within turn and sequence, the accompanying talk, other bodily conduct, the surrounding environment, and the semiotic properties of the gesture—all of which may combine to contextualize the practices accomplished by the gestures at issue.

From the morphological perspective, the addressee-directed pointing gestures were identified by "movement toward" (Eco, 1976, p. 119) the addressee, using an extended finger. Usually, the index-finger morphology was used; however, in rarer cases, pointing at the addressee was accomplished using other fingers, such as the ring finger, while the little finger was also extended (one token), and the little finger (two tokens). These ring-finger and little-finger pointing gestures can be considered as *ad hoc* forms of pointing that in both cases were driven by "biomechanical ease" (*cf.*

1 For examples of such ambiguous cases see Excerpts 1 and 6, line 8.

Cooperrider, 2024). The ring-finger pointing appeared when the speaker held a glass in her hand; therefore, it seems that, from the physiological perspective, the index finger (together with the middle finger) was preferred for holding the glass. The little-finger pointing occurred in an interaction involving four participants, and seemed to be influenced by the participants' seating arrangement. This gesture was produced using the little finger of the right hand when the speaker turned to the participant sitting next to him from the right side. It seems that, in this condition, performing the gesture using the little finger took less effort since the speaker only needed to move the finger slightly to the right; had he pointed using his index finger, he would have had to move his entire hand.

When the index-finger morphology was used, the index finger was clearly protruded more than any other finger (in some cases the middle finger was also extended), with the thumb and other fingers remaining flexed. Flexing and extending varied in degree. Usually, index-finger pointing occurred with pronated orientation, but six times it occurred with supinated orientation. The gesture was produced by pivoting the arm from either the shoulder or from the elbow. Although such morphological differences may have an impact on the interaction between participants (*cf.* Holler, 2010; Kendon, 2004) and reveal further form-function correlations, the wide morphological variety found in the data revealed only few examples of each type, thus making meaningful quantitative comparisons difficult to accomplish.

3 Referential pointing at the addressee

Referential pointing occurs when the reference to the pointed-at participant is part of the propositional content. In such cases, the gesture is often coordinated with an utterance that includes indexing the second person. Such pointing is illustrated in Excerpt 1, which is taken from a conversation between two friends, Sigal and Orly. Prior to the segment shown in Excerpt 1, Orly told Sigal that she would like to go the United States, but her visa had expired.

Ex. (1) 'Visa'

01 Orly: #lax* yesh viza# be-tokef,
to.you there.is visa in-validity
you do have a valid visa,
***Image 1**
#index-finger pointing at Sigal#

Image 1 Orly points at Sigal

02 lo?
NEG
don't you?

FIGURE 1
Excerpt 1

EXCERPT 1

Coordinated with the index-finger pointing at Sigal (Image 1), Orly asks Sigal whether she has a valid visa (lines 1–2). Orly's utterance includes a verbal reference to Sigal in the form of the second person dative pronoun *lax* "to you" (line 1); thus, such pointing can be considered to be referential. However, this is an ambiguous case, since the pointing occurs in the context of eliciting a response, in which the gesture was also found to be used without any verbal reference to the second person, as will be shown in Section 4.

Referential pointing can also occur in cases in which the addressee stands in a metonymic relation of speaker for utterance or speaker for action. As in prototypical pointing a pointed-at object is usually under a particular description (Clark, 2003, p. 247), the content of the addressee's utterance or the action that they perform is often evaluated by a pointer in a verbal component that is co-produced with the pointing gesture.

Excerpt 2 illustrates metonymic pointing in which the addressee represents the content of her previous talk. The example is taken from a conversation between two friends, Naomi and Kelsey. Prior to the segment in Excerpt 2, Kelsey told a story about a couple, friends of hers, who had a big celebration of their marriage proposal as if the proposal was happening in real time. However, the actual proposal had already been made previously and the couple had even set a date for the wedding. Kelsey and Naomi try to understand why they needed to have such an event. Kelsey says that maybe it had something to do with the ring (lines 1, 3); Naomi points at Kelsey with her index finger (Image 2) as she utters *ze mamash muzar* "that is really weird" (line 4), evaluating the story that Kelsey told as being extremely odd. In this case, the *discourse deixis* (Cornish, 2011, 2012) is accomplished not only gesturally, but also verbally by deploying the relative pronoun *ze* "this."

In Excerpt 3, the index-finger pointing occurred when the pointer treated the addressee's way of behaving with ridicule. The example is taken from a conversation involving a couple, Alon and Hillel. Alon is sitting with their baby cradled in his arms. After 37 min of conversation, their friend, Einav, joins them. We enter the interaction after Einav has interacted with Hillel and Alon for about six and a half minutes.

Hillel addresses Alon via a question, also attracting his attention via touch (Image 3), expressing surprise that Einav did not notice [that a camera and a recording device were in the room] (lines 1–5). Alon confirms this via *naxon* "right" (line 6) and Einav asks what it is that she did not notice (line 7). Hillel explains that a camera and a recording device are in the room, and points to these objects (lines 8, 10–11). Einav conveys surprise via *ma?* "what?" (line 13), while Hillel and Alon gaze at her and smile. Hillel and Alon then start laughing, and Hillel points at Einav with his index finger (Image 4). By pointing at Einav, Hillel metonymically spotlights her behavior, locating it as the cause of laughter (*cf.* Yasui, 2023) and evaluating it verbally as great (line 15).

In Excerpts 2 and 3, the pointed-at participants metonymically represented what they had said previously or were currently doing, respectively. Such examples of target-referent metonymy can be characterized as involving a "chain of indicating" (Clark, 2003, p. 264), namely static structures that can be examined link by link, in which the speaker indicates the addressee which, in turn, indexes the referent—utterance or action. However, Cooperrider (2014) argues that, from a cognitive perspective, the target-referent metonymy may be better characterized as being driven by *compression* (Fauconnier and Turner, 2002), which is a feature of conceptual integration. In this

Ex. (2) 'Proposal'

```
01 Kelsey:  ...'aval ha-taba'at,
                but  DEF-ring
            ...but the ring,
02 Naomi:                      ze,
                               this,
03 Kelsey:  ve-ha-ze.
            and-DEF-this
04 Naomi:          #ze   *mamash# muzar,
                    this  really  weird
                   that's really weird,
                         *Image 2
                   #index-finger pointing at Kelsey#
```

Image 2 Naomi points at Kelsey

FIGURE 2
Excerpt 2

case, the pointed-at participants and their actions or utterances are connected via a conceptual relation.

The detailed description and distribution of various types of referential pointing directed at the addressee will be addressed in future research. This paper will now shift its focus to non-referential pointing, which constitutes the central concern of the current study.

4 Non-referential pointing at the addressee

The contexts in which the non-referential index-finger pointing directed at the addressee were employed appeared to be diverse. In the majority of the cases ($N = 41$), the gesture occurred in contexts that shared a broad sense of opposition: some were disaffiliative contexts associated with dispreferred actions, such as disagreement, disconfirmation, and repair, while in others speakers conveyed information that was (assumed to be) contrary to the addresses' expectations. In these contexts, speakers typically established and maintained a convergent status and stance of *epistemic authority* (Heritage, 2012). In other contexts, recipients displayed cues of disengagement in the interaction ($N = 15$) or speakers attempted to elicit a (minimal) response from the addressee ($N = 13$). Additionally, some occurrences followed the completion of a cognitive process ($N = 7$), typically related to remembering, while others were associated with interruptions and discourse suspension ($N = 5$). Sometimes it was challenging to delineate among these categories, as some occurrences fit into more than one category.[2] In what follows, I will illustrate these contexts.

4.1 Disconfirmation

In Excerpt 4, the gesture is associated with disconfirmation. Prior to the excerpt, Dotan told Alex that he had attended a concert held at the singer's house. After Dotan explained exactly where the concert took place, Alex asks whether everyone there is a musician (line 1). Dotan disconfirms this via *lo* "no" (line 2), stating that at least one hairdresser lives there (line 3), while pointing at Alex with his index finger (Image 5). He later explains (not shown) that this knowledge was based on the fact that his wife once visited that hairdresser.

4.2 Disagreement

In Excerpt 5, the gesture is associated with disagreement. The excerpt is taken from a conversation between two friends, Dov and Boaz, which revolves around politics and the controversial status of the current Israeli Prime Minister. In Excerpt 5, Boaz disagrees with Dov's definition of democracy and then explains why.

Dov claims that the tenure of the current Israeli prime minister is a consequence of democracy, which is the will of the people (lines 1–3). In response to Dov's definition of democracy as the will of the people (line 2), overlapping with Dan, Boaz expresses disagreement via headshakes (line 3). Then, prefaced by a click (line 4) expressing a negative stance (Ben-Moshe and Maschler, Forthcoming), Boaz proceeds to express disagreement verbally, saying that democracy is more complex than the will of the people (line 5). This utterance is associated with two pragmatic gestures: Boaz first briefly points at Dov with his index finger (Image 6) and then performs the Palm Up Open Hand gesture which is often used to frame a content as obvious, self-evident, or as shared knowledge (e.g., Inbar and Maschler, 2023). Dov requests explanation (line 7), thus admitting Boaz's *epistemic authority* (Heritage, 2012) on the subject. In response, Boaz provides a list of explanations (lines 9–13), coordinated with a particular type of listing gesture—the Finger-counting gesture (Images 7 and 8), which tends to appear in contexts of opposition and is associated with epistemic authority among Hebrew speakers (Inbar, 2020; Inbar, Forthcoming).

4.3 Displaying epistemic authority

Interestingly, a demonstration of knowledge was observable in most of the cases in which the index-finger pointing directed at the addressee occurred in disaffiliative contexts or when the addressee did not expect the information that was provided. This was the case in Excerpts 4 and 5. In Excerpt 4, Dotan was the person who knew who lived in the place that was being discussed, and Alex addressed him by asking a question regarding this issue. In Excerpt 5, Boaz positioned himself as being more knowledgeable about politics when he disapproved of Dov's definition of democracy, and Dov addressed Boaz with the question "why?" Moreover, Boaz deployed the Finger-counting gesture, which is another strategy that is found in contexts in which speakers produce and maintain convergent status and stance of epistemic authority (Inbar, 2020, 2024).

2 In such cases, the assignment to a category was based on the salient features that were reflected in other gestural or verbal means. Thus, the distribution presented above is provided to offer a general impression of the data and should not be interpreted as a precise representation.

Ex. (3) 'Camera'

```
01 Hillel:  $'eyze ^mag$niv ze  'aval?*$
              which cool    this but
              how cool is it though?
                                        *Image 3
            $.........$touches Alon with index finger$
                  ^gazes at Alon----->l. 7
   alon:    >>&gazes at baby---------->l. 7

02          ...kolet?
               take.in.PRS.SGM
            ...[do you] realize [that]?

03          ke'ilu,
            like,

04          hi  nixnesa,
            she enter.PST.3SGF
            she entered,
05          lo sama.lev      'aval.
            NEG notice.PST.3SGF but
            but didn't notice.

06 Alon:    ...naxon.
            ...right.

07 Einav:   she^-ma?&
            that-what?
            what?

08 Hillel:  ...ψshe-yesh      po   maxshir haklata,ψ
               that-there.is  here device  recording
            ...that there is a recording device,
               Ψpoints at recording device-------->ψ

09          ±'al ha-shulxan,
              on DEF-table
              on the table,
   alon:    ±gazes at Einav, smiles--->l.14

10          ϕve,
             and,
            ϕpoints at camera--->l.11

11          matsle÷ma,ϕ
            a camera,
                  ÷gazes at Einav, smiles--->l.16

12          she-metsalemet  'otanu.
            that.film.PRS.SGF us
            that is filming us.

13 Einav:   (0.8) ma?
            (0.8) what?

14 Alon:    ...@±

15 Hillel:  #¶(1.0) me'ule!*#
              (1.0) great!
                              *Image 4
             ¶laughs--------->>
           #index-finger pointing at Einav#

16 Einav:   @÷

17          ma   kore?
            what happen.PRS.SGM
            what's going on?
```

Image 3 Hillel touches Alon

Image 4 Hillel points at Einav

FIGURE 3
Excerpt 3

More striking examples of a display of epistemic authority are illustrated in Excerpt 6, which is taken from a conversation between two friends, Kelsey and Naomi. The participants discuss the use of disposable utensils. Prior to what is shown in Excerpt 6, Kelsey said that such use was unreasonable, but that she could understand her mother using disposable utensils when having many guests. She adds that in such cases, her mother should use disposable utensils made of paper (lines 1–3). In what follows, Naomi, who is currently attending a textile school, explains that most disposable utensils made of paper cannot decompose or be recycled.

In response to Kelsey's statement that it is better to use disposable utensils made of paper (lines 1–3), Naomi produces a click (line 6), projecting disaffiliation, and she will later explain that such material cannot actually decompose or be recycled (lines

Ex. (4) 'Hairdresser'

01 Alex: ...kulam sham muzikaim?
everyone there musicians
...**everyone there is a musician?**
{----------*sf*---------}

02 Dotan: ....lo,
NEG
....**no,**

03 ...#yesh sham lefaxo*t# saparit 'axat.
there.is there at_least hairdresser.SGF one.F
...**there is at least one hairdresser there.**
*Image 5
#index-finger pointing at Alex#

Image 5 Dotan points at Alex

FIGURE 4
Excerpt 4

22–24). Prefaced by two discourse markers—*gam kaxa* (lit.) "also so" conveying concession or returning to an earlier subject after a digression, and *rak she-tid'i* "just so you know/FYI" conveying epistemic authority (line 8), Naomi raises the issue of disposable utensils made of paper for further discussion (lines 8–12). When she utters *gam kaxa rak she-tid'i she-ba-niyar*, "anyhow just so you know that [those made] of paper," she points at Kelsey with her index-finger (Image 9). Although the verbal form *tid'i* consists of the reference to the second person, the gesture occurs in the opposing context and is associated with epistemic authority by virtue of being co-produced with such a phrase. Naomi then projects a violation of expectation via *be'etsem* "actually" (line 14) and begins to explain about the materials of which such utensils are made (lines 14–15), but encounters a problem retrieving the exact term. After a pause of 0.5 s, she provides a general formulation *xomer* "material" (line 16) and, after an additional 0.6 s, Kelsey suggests a potential candidate "plastic" (line 17), which Naomi confirms via nodding (line 18). After an additional hesitation (line 19), Naomi finally retrieves the professional term of the material of which such disposable utensils are made—polyurethane (line 20)—followed by "I think" to mark uncertainty (Ziv, 2016). Naomi then notes that she has learned about it at (textile) school (line 21), pointing at Kelsey again using her index finger (Image 10). By indicating where this knowledge was acquired, Naomi establishes her credibility and expertise on the subject. The utterance co-produced with the gesture does not include any reference to the second person, and the gesture appears to be employed to display epistemic authority. This epistemic status could be weakened by the speaker's hesitations and expressions of uncertainty, making its reinforcement by the gesture particularly relevant in this moment of the interaction. Moreover, as in other occurrences, epistemic authority is further conveyed within a broad context of disaffiliation, where Naomi challenges Kelsey's assertion that disposable utensils made of paper are preferable. Naomi then concludes that such material cannot decompose or be recycled (lines 22–24), coordinated with the Raised Index Finger gesture (Image 11), which is another means found to be associated with epistemic authority (Inbar, 2022).

In Excerpt 7, which is a continuation of the conversation between Kelsey and Naomi in which they discuss the use of disposable utensils, pointing at the addressee occurs five times. Naomi initiates a new telling regarding allegedly disposable utensils (lines 1–7). She first introduces this topic by stating that there are disposable utensils that are marketed as being degradable in compost (lines 5–7), ending with "continuing appeal intonation" (Du Bois, 2012, 5.3), which is also characterized as "try-marked" intonation (Sacks and Schegloff, 1979), an intonation contour which, in Hebrew, is designed to prompt a (minimal) response from the listener while signaling that there is more to be conveyed. Kelsey responds with *naxon* "right" (line 8), confirming that she is familiar with such utensils, upgrading her epistemic certainty by adding that she is also familiar with bags made from this material (line 9). Overlapping with Kelsey, Naomi projects surprise and unexpectedness by both uttering *ve* "and" (*cf.* Hopper, 2021) and employing a co-produced Raised Index Finger gesture (Image 14; Inbar, 2022), which she then transforms into a pointing gesture directed at Kelsey (Image 15) while strongly objecting to the information by stating *ze lo* "they are not" (line 11). Kelsey responds with her mouth wide open (Image 16), which is considered to be one of the components of a facial surprise display (e.g., Darwin, 1998; Reisenzein et al., 2012). Naomi then begins to explain how this substance can decompose (line 13), reproducing the index-pointing gesture directed at Kelsey to display epistemic authority. However, Naomi then hesitates and averts her gaze (line 14; Image 17), displaying a "thinking face" (e.g., Goodwin and Goodwin, 1986; Bavelas and Chovil, 2018). Interestingly, Naomi withdraws the index-finger pointing gesture as she starts to hesitate and reproduces it when she proceeds with her explanation (lines 15–16) (see discussion below on indicating the accomplishment of a cognitive process). Kelsey responds again with her mouth wide open (line 17). Naomi then tells Kelsey about her personal experience regarding such utensils, namely, that her mother put them into the regular compost, and they did not decompose (lines 20–21), deploying the index-finger pointing once again. By saying this, Naomi reinforces her epistemic authority by adding personal experience to her theoretical knowledge.

4.4 Gaining attention

In face-to-face interaction, attention can be signaled by a gaze directed at the speaker (e.g., Clark and Brennan, 1991). The findings suggest that after being disrupted, the interlocutor's attention can be regained by the speaker pointing at their addressees. The current study revealed 15 instances of such addressee-directed pointing. The first occurrence of index-finger pointing in Excerpt 7 (Image 13) can be viewed as an example of the phenomenon. At the beginning of the excerpt, Naomi initiates a new telling regarding a particular kind of disposable utensil, prefacing her telling with a cluster of pragmatic markers (lines 1–3). During this prefacing, the recipient, Kelsey, averts her gaze (line 3) and starts to examine her fingers (Image 12). This type of behavior reveals a degree of reduction in her engagement in the interaction. It is plausible that, at this moment during the interaction, Naomi uses the pointing gesture as a device to attract Kelsey's attention and to ensure her involvement. In other words, by pointing at Kelsey, Naomi could signal that she needs positive evidence of her attention. In fact, after

Ex. (5) 'Democracy'

01 Dov: ratsita demokratia?
want.PST.2SGM democracy
didn't you want democracy?

02 ....ze retson ha-'am.
this will.of DEF-people.
....this is the will of the people.

03 %...%hem rotsim 'oto,
they want.PRS.PL him
...they want him,

boaz: %headshakes%

04 Boaz: **tsk**

05 demo#kratia* ze# yoter murkav ^mi-retson ha-'am.^
democracy this more complex from-will.of DEF-people
democracy is more complex than the will of the people.
*Image 6
#index-finger pointing at Dov#
^PUOH---------->^

06 Dov: 'afilu 'im hu shakran.
even if he liar
even if he is a liar.

Image 6 Boaz points at Dov

07 lama?
why?

08 Boaz: **tsk**

09 ±ki be-demokratia yesh hagana 'al mi'uti--*m,
because in-democracy there.is protection on minorities
because in democracy there is protection for minorities,
*Image 7
±Listing Gesture ----------> 1.13

10 yesh xuki--*m,
there.are rules
there are rules,
*Image 8

11 mesuyamim,
certain
certain [rules],

Image 7

Image 8

Image 7–8 Boaz produces Listing Gesture

12 she-'ata yaxol leshanot 'otam be-'ofen demokrati,
that-you can to.change them in-way democratic
that you can change democratically,

13 'aval yesh xukim mesuyamim,
but there.are rules certain,
but there are certain rules,
(Listing Gesture)---------->>

FIGURE 5
Excerpt 5

EXCERPT 5

Naomi points at her, Kelsey directs her gaze at Naomi again (line 5). Note that the context in which the gesture occurred cannot be characterized as opposition, nor is it one of the other specific contexts identified in this study in which index-finger pointing gestures occurred, except for those in which the interlocutor's attention was diverted.

Excerpt 8, which is taken from a conversation between two friends, Lital and Eden, is another instance of using such pointing to

Ex. (6) 'Disposable utensils'

01 Kelsey: lefaxot,
at least,

02 ki'ilu she-tishtadel she-ze yihye niyar,
like that-try.FUT.3SGF that-this be.FUT.3SGM paper
like she should make an effort it would be the paper

03 ve-lo plastic.
and-NEG plastic
and not plastic.

04 lo yoda'at.
NEG know.PRS.SGF
[I] don't know.

05 ze lo sababa.
this NEG cool
it's not cool.

06 Naomi: tsk

07 Kelsey: xad.pa'ami ze /ro'a/.
one-time this /evil/
disposable utensils are evil.

08 Naomi: #gam kaxa rak she-tid'i she-ba-niya*r,#
also so just that-know.FUT.2SGF that-in.DEF-paper
anyhow just so you know that [those made] of paper,
*Image 9
#index-finger pointing at Kelsey---------->#

Image 9 Naomi points at Kelsey (a)

09 kol ha,
all DEF
all the,

10 xadapim,
disposable utensils,

11 ...'ulay lo kol,
maybe NEG all,
...maybe not all [of them],

12 'aval harbe me-ha-xadapim she-hem niyar,
but many from-DEF-disposable.utensils that-they paper
but many of the disposable utensils that are made of paper,

13 Kelsey: ...**mhm.**

14 Naomi: 'az ze be'etsem,
so this actually
so it's actually,

15 shxavot shel,
layers of,

16 (0.5)xomer,
(0.5)material,

FIGURE 6a
Excerpt 6 (lines 1–16)

7 Kelsey: (0.6)plasti?
(0.6)plastic?

8 Naomi: **(nods)**

9 shel eh,
of eh,

0 poli'orten 'ani xoshevet,
polyurethane I think.PRS.SGF
polyurethane I think,

1 #'a*naxnu lamadnu 'et ze ba-limudim.#
we learn.PST.1PL ACC this in.DEF-studies
we learned it in school.
*Image 10
#index-finger pointing at Kelsey#

Image 10 Naomi points at Kelsey (b)

2 (0.6) &ve-'az,
(0.6) and so,
&Raised Index Fingers----->1.24

3 ...'i 'efshar lemaxzer 'et ze.
NEG possible recycle.INF ACC this
... it cannot be recycled.

4 ...ve-ze gam lo mit*kale.&
and-this also NEG decompose.PRS.SGM
...neither can it decompose.
*Image 11

Image 11 Naomi raises index fingers

FIGURE 6b
Excerpt 6 (lines 17–24)

attract the addressee's attention to the upcoming talk. After a long pause (five seconds) during which there was no interaction between Eden and Lital—Eden was gazing at her cellphone and Lital was looking away—Lital initiates a turn saying *'az* "so" (line 1), in continuing intonation, but Eden interrupts Lital by launching a course of action (*cf.* Sidnell, 2007) uttering *takshivi*! "listen!" (line 2), overlapping with Lital. The Hebrew verb *takshivi* "listen" can be considered as an attention-drawing device (e.g., Aijmer, 2010) and is used here in exclamatory intonation. Using *extreme case formulation* (Pomerantz, 1986), Eden then tells Lital that everyone is talking in the same way as she is (*kulam medabrim, kamoni!*), again using an exclamatory intonation (lines 3–4). Eden turns her gaze at Lital and Lital at Eden (line 4). Eden then produces the particle *hine* (line 5) which, in this case, indicates that Eden has visual access to the entity that serves as evidence for her previous statement (Shor et al., Forthcoming; Shor et al., Forthcoming), and points at Lital with her index finger (Image 18). Via this pointing, Eden obtains Lital's attention toward her upcoming talk—reading her friend Gil's WhatsApp message. Eden then points at her cellphone screen (Image 19), even though Lital cannot see the screen. While pointing at her cellphone screen, Eden starts reading the message in which Gil used the formulaic expression *'avarti shinuy, 'ani kvar* X "I've changed, I'm already X" (lines 8–11) that Eden had been using frequently, as she later explains (lines 13–15).

4.5 Indicating accomplishment of cognitive process

Index-finger pointing was also observed in contexts where the speaker had just undergone a certain cognitive process, often related to remembering. Conversational remembering has been claimed to be a systematic and joint activity that is performed for interactional purposes (e.g., Carranza, 2016; Hirst and Echterhoff, 2012; Middleton and Edwards, 1990). Carranza (2016) described three types of remembering sequences—*assisted*, *metacognitive*, and *spontaneous remembering*—all of which were attested in the current study as being associated with the index-finger pointing gesture directed at the addressee. One instance of index-finger pointing occurred in a *metacognitive remembering* sequence in which the speaker achieved remembering via the metacognitive strategy of a reflective question addressed to herself ("What else did she tell me?"). Another instance

Ex. (7) 'Non-biodegradable Material'

```
01 Naomi:   ve,
            and,
02          bli     kesher,
            without connection
            anyways,
03          ...ψratsiti     lehagid she*,
                want.PST.1SG to.say  that,
            ...I wanted to say that,
                                    *Image 12
   kelsey:     ψgaze aversion; gazes at hands-->1. 4
```

Image 12 Kelsey gazes at hands

```
04          (0.7)#yesh*ψ,#
                  there.are
            (0.7)there are,
                    *Image 13
                 #index-finger pointing at Kelsey#
```

Image 13 Naomi points at Kelsey (a)

```
05          &(0.5)kelim   had-pa'amiyim she-meshavkim    'otam ke,
                utensils disposable    that-market.PRS.PL them like,
             (0.5)disposable utensils that are marketed as,
   kelsey:  &gazes at Naomi----------------------------------->1.20
06          (0.7)she-'efshar  lasim 'otam  be-kompost
                 that-possible to.put  them  in-compost
            (0.7)you can put them in the compost
07          ve-hem  mitkalim?,
            and-they decompose.PRS.PL
            and they decompose?,
```

FIGURE 7a
Excerpt 7 (lines 1–7)

occurred in an *assisted remembering* sequence, in which a reminder was provided by the other participant. Yasui (2017) found similar cases in Japanese interaction, suggesting that the addressee-directed gestures may indicate that the source of the remembering was contained within what the current or prior speaker had just said, and were thus used as "touched-off" markers (Jefferson, 1987; Sacks, 1992). Finally, five remaining instances occurred in *spontaneous remembering* sequences, in which there were no visible reminders; that is, there was no talk that was designed to elicit remembering. Such cases are illustrated in the following two examples.

Excerpt 9 is taken from an interaction between two friends, Amit and Tom, who work as waitresses at the same place. The conversation revolves around a joint work shift. Amit begins a narrative about what happened on Thursday with their co-worker Noam, who was nervous and angry that day (lines 1–2, 5–7), and mentions that Tom was working on the same day (lines 3–4). After a stretch of talk (not shown) in which Amit recounts the chain of events during that shift, Tom suspends the discourse by noting that she does not remember Noam looking nervous (line 9). In overlap with the end of Tom's utterance, Amit produces a click (line 10), projecting disaffiliation (Ben-Moshe and Maschler, Forthcoming). She then confirms that Noam was nervous, reinforcing her statement via the intensifier *mamash* "really" (line 11). Amit attempts to proceed with her talk (line 12), but Tom interrupts her again by upgrading her previous statement and saying that she actually does not remember anything about that shift (line 13). She then begins to recall what happened during that shift (line 14), but Amit produces another click (line 15), conveying impatience (Ben-Moshe and Maschler, Forthcoming), and again attempts to proceed with her talk (lines 16–17, 19). However, Tom interrupts Amit again, deploying the *change-of-state marker* (*cf.* Heritage, 1984, 2016) *ah* (line 18) to indicate that she has undergone a cognitive process, followed by spontaneously conveying remembering (line 20), overlapping Amit's talk. The display of remembering is coordinated with pointing at Amit (Image 20). Amit reproduces her utterance for the fourth time, and finally manages to complete it (lines 21–22).

Another occurrence is illustrated in Excerpt 10, taken from an interaction in which Dotan attempts to explain how to get to a certain location in the city. After Dotan's unsuccessful attempts to describe the exact location (not shown), Alex concludes that it is somewhere in the city center (lines 1–2). Dotan objects (lines 3–4) and, using a continuing appeal intonation, offers additional coordinates (lines 6–7), which Alex confirms via "Ok" (line 8) after a long pause. After another long pause (1.6), an inhalation, and the hesitation marker *e--hm* (lines 9–10), Dotan deploys the change-of-state marker *ah!* (line 11) to register either that he has undergone the

08 Kelsey: ...naxon.
...right.

09 ...gam saklyot ka'ele.
also packets like.these
...[there are] also bags like these.

10 Naomi: ±ve*,±
and,
±Raised Index Finger±
*Image 14

Image 14 Naomi performs the Raised Index Finger gesture

11 #ze lo*.
this NEG
they are not.
*Image 15
#index-finger pointing at Kelsey-->l. 12

Image 15 Naomi points at Kelsey (b)

12 Kelsey: **ϕ*opens mouth-->l. 14**
(0.5)#
***Image 16**

13 Naomi: (0.5)#ze mitkale--,
this decompose.PRS.SGM
(0.5) they decompose,
#index-finger pointing at Kelsey--->l. 14

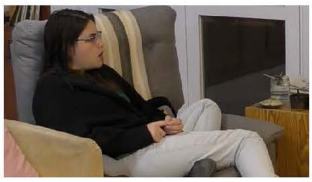


Image 16
Kelsey opens mouth

14 ...#¥be--*,Φ
...i--n,
¥gaze aversion--->l. 15
***Image 17**

15 (0.5)¥◊#kompost ta'asiyati,
compost industrial
(0.5) industrial compost,
#index-finger pointing at Kelsey-->l. 16
◊gazes at Kelsey--->l. 18

Image 17
Naomi's gaze aversion

16 ...she-samim be-ze kol-mine xomarim
that-put.PRS.PLM in-this every-sorts.of materials
...that they put all kinds of substances in
she-gormim le-ze lehitparek,#
that-cause.PRS.PLM to-this to.break.down
that cause it to break down,

17 Kelsey: **ϕ*opens mouth-->l. 19**

18 Naomi: ◊ve,
and,

19 nagid,Φ
say.FUT.1PL
let's say,

20 (0.8)#'ima sheli sama ba-kompost& ragil?,
mother mine put.PRS.SGF in.DEF-compost regular
(0.8) my mom put [it] in regular compost?,
#index-finger pointing at Kelsey---------->l. 21

21 (0.7)#ze lo be'emet matsliax.
this NEG really succeed.PRS.SGM
(0.7) it doesn't really work.

FIGURE 7b
Excerpt 7 (lines 8–21)

cognitive process of remembering or that he has just figured out what is the best way to describe the direction. Dotan then produces the audible gesture of snapping his fingers (line 12), which, in various cultures, can be used to attract attention (e.g., Bowles, 2017; Christidou, 2018; Will, 2021). He then produces another hesitation marker *ne--hm* (line 13) coordinated with the index-finger pointing at Alex (Image 21), followed by offering another coordinate—the tunnel near Gan Sacher (lines 14–17).

In such contexts, the gesture was often produced after pauses and hesitation markers on the part of the current speaker. Therefore, it is conceivable that the gesture was used to attract the interlocutor's attention after the interlocutor could potentially have been distracted by such disfluencies, similar to the cases illustrated in Excerpts 7 and 8. In other cases, calling for attention to their remembering, speakers interrupted their co-participants, as in Excerpt 9. However, some occurrences of index-finger pointing at

Ex. (8) 'I've changed'

```
01 Lital:   (5)'az,
               so,
02 Eden:       takshivi!
               listen.FUT.2SGF
               listen!
           >>gazes at cellphone, smiles--> l. 4(^)
03         kulam    medabrim,
           everyone talk.PRS.PL
           everyone is talking,
04         kamo^&$ni!
           like.me
           like me!
                &gazes at Lital--> l. 6
   lital:        $gazes at Eden--> l. 7
05         #*hine,#
             PART
             here,
            *Image 18
           #index-finger pointing at Lital#
06         ±ya&
           ±points at cellphone screen---> l. 7
07         yadid$  sheli 'axshav ko*±tev      li,
           friend  mine  now     write.PRS.SGM to.me
           my friend is texting me now,
                                   *Image 19
08         day eden,
           PART Eden
           enough Eden,
           {reading}
09         'ani lo  ba         maxar,
           I    NEG come.PRS.SGM tomorrow
           I'm not coming tomorrow.
           {reading}
10         'avarti       shinuy,
           undergo.PST.1SG change
           I've changed,
           {reading}
11         'ani kvar    lo gil shel ha-tsxokim.
           I    already NEG Gil of   DEF-laughs
           I'm no longer the Gil who laughs a lot.
           {reading}
12         (0.5)
           (gazes at Lital, smiles)
13         ki     kol ha-zman 'ani 'osa      lahem,
           because all DEF-time  I    do.PRS.SGF to.them
           because every time I tell them,
14         'avarti       shinuy,
           undergo.PST.1SG change
           I've changed,
15         'ani kvar    meshuxreret.
           I    already released
           I am already carefree.
```

Image 18 Eden points at Lital

Image 19 Eden points at cellphone

FIGURE 8
Excerpt 8

the addressee were associated with interruption and discourse suspension in other contexts, not necessarily those in which remembering was displayed (see Excerpt 11).

4.6 Interruption/discourse suspension

Another example of discourse suspension is illustrated in Excerpt 11, taken from a conversation between two acquaintances, Sara and Dina. Sara is a resident of the small settlement where Dina works. Sara begins to tell Dina about how she manages the treatment of her sick daughter. She says that she stays near her daughter constantly (line 1), and then starts to explain why (lines 2–3), but Dina's cellphone suddenly rings.

As her cellphone rings, Dina interrupts Sara by deploying two discourse markers of *suspension* (e.g., Scott, 2002), *slixa* (*cf.* Inbar, 2022), and *rak rega* (*cf.* Bardenstein and Shor, 2019), which serve to indicate an attempt to stop others from speaking in order to gain or keep the floor. These discourse markers are coordinated with index-finger pointing at Sara (Image 22).

Ex. (9) 'Work shift'

01 Amit: kitser,
short
in short,

02 yom xamishi no'am hayta
day fifth noam be.PST.3SGF
on Thursday Noam was

03 'at 'avadt,
you.SGF work.PST.2SGF
you were working,

04 leda'ati.
to-my.knowledge
in my opinion.

05 ...hi hayta--,
she be.PST.3SGF
...she wa--s,

06 ...'atsbanit ve--,
...nervous a--nd,

07 ve-ko'eset,
and angry,

((86 intonation units omitted))

08 Amit: ve-'az eh,
and-then eh
and then eh,

09 Tom: 'ani lo zoxeret she-hi nir'eta li 'atsbanit.
I NEG remember.PRS.SGF that-she looked to.me nervous
I don't remember her looking nervous.

10 Amit: **tsk**

11 hi hayta mamash.
she be.PST.3SGF really
she definitely was.

12 ve-'az e--h,
and-then eh
and then e--h,

13 Tom: 'ani lo zoxeret @klum /'aval 'al/
I NEG remember.PRS.SGF nothing /but about/
I don't remember anything /but about/

ha-mishmeret ha-zot,
DEF-shift DEF-this
this shift,

14 'avadnu layla--,
work.PST.1PL night
we worked a night [shift],

15 Amit: **tsk**

16 ve-'az,
and then,

17 yom shishi,
day sixth
[On] Friday,

18 Tom: ah,
NON-LEX
oh,

19 Amit: 'asinu,
do.PST.1PL
we did,

20 Tom: #haya xala*sh.#
be.PST.3SGM weak
it was off-peak.

***Image 20**
#index-finger pointing at Amit#

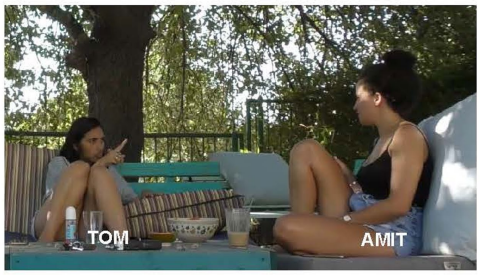


Image 20 Tom points at Amit

21 Amit: ve-'az yom shishi,
and-then day sixth
and then on Friday,

22 'asinu ptixa.
do.PST.1PL opening
we were the openers.

FIGURE 9
Excerpt 9

4.7 Eliciting a (minimal) response

The fact that pointing can be used when a speaker is attempting to elicit a response has been attested in conversation analytic studies (e.g., Bavelas et al., 1995), as well as in experimental studies (e.g., Holler, 2010). The data manifested numerous examples of ambiguous cases in which the gesture was employed in the context of eliciting a response and the utterance coordinated with the pointing included indexing the second person. However, the study revealed 13 occurrences of pointing at the addressee in the context of eliciting a response, in which the pointing was non-referential. Two such occurrences are illustrated in Excerpt 12. The excerpt is taken from a conversation between two friends, Orly and Sigal. Orly's daughter is supposed to celebrate Bat Mitzvah—turning 12 years old, a landmark

Ex. (10) 'Address'

01 Alex: okey,
ok,

02 ..merkaz ha-'ir 'eyfo she-hu.
center DEF-city where that-he
..in the city center somewhere.

03 Dotan: ...lo lo.
NEG NEG
...no no.

04 lo.
NEG
no.

05 ...e--hm
NON-LEX

06 ...beyn betsal'el,
between Bezalel
...between Bezalel[street],

07 ..le-rexov 'aza?,
to-street Gaza
..and Gaza street?,

08 Alex: (1.2)okey [key].
(1.2)ok.

09 Dotan: **(1.6)(inhalation)**

10 **/mts/e--hm**

11 ah!
NON-LEX
oh!

12 **...(snapping fingers)**

13 #...ne--hm*#
NON-LEX
*Image 21
#index-finger pointing at Alex#

Image 21 Dotan points at Alex

14 ha-minhara mitaxat le--em
DEF-tunnel under to--em
the tunnel under e--hm

15 **...mts (inhalation)**

16 ...leyad gan saker?,
near garden sacher
...near Gan Sacher?,

FIGURE 10
Excerpt 10

Ex. (11) 'Phone call'

01 Sara: 'ani tsmuda 'eleya,
I linked to.her
I'm near her,

02 ...lama?
...why?

03 ...⊙ψki 'im 'ani 'ivater,
because if I give.up.FUT.1SG
...because if I give up,
⊙{Dina's cellphone rings->l. 6}
dina: ψgazes at cellphone----->l. 4

04 lo ψ¥yihye
NEG be.FUT.3MSG
there will be no
dina: ¥gazes at Sara--->l. 6

05 Dina: #slixa,
forgiveness
hang on,
#index-finger pointing at Sara->l.6

06 ◊rak *rega,#¥⊙
just moment
just a moment,
*Image 22
◊picks up the phone---->

Image 22 Dina points at Sara

07 ha-xatula sheli,
DEF-cat mine,
my cat,

08 'asta hayom 'ikur.
do.PST.3SGF today sterilization
was spayed today,

09 ...'az,
...so,

10 ze ha-shomer.
this DEF-keeper
it's the [cat] keeper [who is calling].

FIGURE 11
Excerpt 11

in Jewish tradition—and all the family is going to stay in a desert for this celebration.

Coordinated with pointing at Orly with her index finger (Image 23), Sigal requests confirmation (*cf.* Ben-Moshe and Maschler, 2024) that the Bat Mitzvah of Orly's daughter will take place in a month (line 5). Orly disconfirms saying that the celebration will take place in August (line 11). [Since the conversation was recorded in the beginning of June, it turns out that the celebration will take place later than was expected by Sigal.] Co-produced with another pointing at Orly (Image 24), Sigal deploys the change-of-state marker *ah* (line 12), followed by eliciting a minimal response from Orly deploying another confirmation request, this time framing the information requested for confirmation as unexpected via *rak* lit. "only" (line 13). Orly confirms via *ken* "yes" (line 14).

While in some examples of this variety, eliciting a response appeared to be combined with other contextual categories revealed in this study, in three cases, the gesture could also be interpreted as a device used to specify the addressee of an utterance in order to elicit their response (e.g., Bavelas et al., 1995; Enfield et al., 2007). These cases were observed only in interactions involving more than two participants, suggesting that the gesture is a versatile tool that adapts its functions to the dynamics of the interaction. Such an example is shown in Excerpt 13, which is taken from a conversation held during a family meal at Yair and Neta's place with Yair's parents. Prior to what is shown in Excerpt 13, the conversation revolved around salted fish which Yair's parents do not like. Suddenly, Yair recalls a shared experience with his parents about their trip to Jordan where they were stuck without food, and someone brought them a canned fish.

Yair begins to recall a family trip to Jordan (lines 1–2). Then, Yair's mother adds to the previous discussion concerning the salted fish, that smoked salmon is indeed tasty (line 3). Overlapping his mother, using an appeal intonation, Yair addresses his father with the request for confirmation that they had been traveling in the Dana Reserve (lines 4–5) during their trip to Jordan. This confirmation request is coordinated with the index-finger pointing at Yair's father (Image 25) which is held until the information is confirmed via *ken* "yes" (line 6). In what follows, Yair and his father try to bring up memories of that experience whereas the family was stuck without food, and someone

Ex. (12) 'Bat Mitzvah'

01 Sigal: matay ha-bat
when DEF-bat
when [is] the Bat

02 ha-bar mi
DEF.bar mi
the Bar Mi

03 bat mitsva?
Bat Mitzvah?

04 Orly: shel meravi?
of Meirav
Meirav's?

05 Sigal: #'od* hodesh?
more month
in a month?
*Image 23
#index-finger pointing at Orly--->1. 6

Image 23 Sigal points at Orly

06 matay 'atem# nos'im?
when you.PL go.PRS.PL
when are you going?

07 Orly: **(1.3)/e m/**

08 Sigal: hizmant 'et ha
book.PST.2SGF ACC DEF
did you book the

09 Orly: be-'ogust.
in-august
in August.

10 Sigal: #ta*h,t#
*Image 24
tnodt
#index-finger pointing at Orly#

Image 24 Sigal points at Orly

11 ze be-'ogust rak?
this in-august only
[will] it [be] then in August?

12 Orly: (0.7)ken.
(0.7)yes.

FIGURE 12
Excerpt 12

Ex. (13) 'Salted fish'

01 Yair: 'ani zoxer pa'am 'axat,
I remember.PRS.MSG time one
I remember once,

02 be-yarden.
in-jordan
in Jordan.

03 Mom: salmon me'ushan ze davar ta'im.
salmon smoked this thing tasty
smoked salmon is a tasty thing.

04 #kshe,
when,
#index-finger pointing at Dad--->1. 6

05 tiyalnu be-dana*?
travel.PST.1PL in-dana
we traveled in the Dana [Reserve]?
*Image 25

Image 25 Yair points at Dad

06 Dad: ...ke--n.#
...**ye--s.**

07 Yair: ...mish

08 lo haya me-'eifo,
NEF be.PST.3SGM from-where
there was no from where,

09 lama,
why,

10 Dad: haya 'iru'a 'az,
be.PST.3SGM event then
there was an event then,

11 Yair: 'aval haya,
but be.PST.3SGM
but [there] was,

12 Dad: ken,
yes,

13 ken,
yes,

14 Yair: haya mashehu,
be.PST.3SGM something
there was something,

15 Dad: lo haya lanu klum!
NEG be.PST.3SGM to.us nothing
we had nothing!

FIGURE 13
Excerpt 13

brought leftovers from some event (lines 7–15). In this example, the pointing occurs in the context of eliciting a response from Yair's father, but since this interaction is multi-party, the gesture could also be interpreted as a device used to specify the addressee.

5 Discussion

The study showed that non-referential pointing directed at the addressee occurred in various contexts. In most of the cases ($N = 41$), the gesture occurred in contexts involving an action that did not support or endorse the co-participant's stance or point of view. Some of them were disaffiliative contexts associated with dispreferred actions, such as disagreement, disconfirmation, and repair, while in others, the speakers conveyed information that was (assumed to be) contrary to the addresses' expectations. In these contexts, the speakers typically produced and maintained a convergent status and stance of epistemic authority. Additionally, the gesture occurred when the addressee conveyed cues of lack of engagement in the interaction ($N = 15$), when they were indicating accomplishment of a cognitive process ($N = 7$), when the speakers attempted to elicit a (minimal) response from the addressee ($N = 13$), and when discourse suspension or interruption occurred ($N = 5$). The analysis of the examples shows that these contexts occasionally had an overlapping nature, making delineation among them challenging.

The pertinent question that emerges is why the gesture in question appears across these various contexts. It can be assumed that, similar to prototypical pointing, usually defined as a bodily movement toward a target in order to direct someone's attention to it (e.g., Clark, 2003; Cooperrider, 2023; Cooperrider et al., 2018; Eco, 1976; Kendon, 2004; Kita, 2003), the non-referential index-finger pointing gesture directed at the addressee may primarily function as a mechanism for capturing attention. However, while the prototypical function that is commonly attributed to pointing gestures entails redirecting a listener's attention to a referent (e.g., Cooperrider, 2023) that is most often presumed to be visual and present in the speech situation, the index-finger pointing at the addressee that was examined in this study mainly serves to attract the addressee's attention to the pointer's upcoming utterance. Interestingly, research

in experimental psychology has shown that recipients are most likely to notice deictic gestures while interpreting speech (e.g., Langton and Bruce, 2000). The heightened awareness of these gestures suggests that they may play a crucial role in directing the recipient's focus, thereby reinforcing the hypothesis that these gestures function as an attention-drawing device.

Existing research has revealed various motivations for capturing interlocutor's attention, with this function being assigned to a range of devices (e.g., Aijmer, 2010; Atkinson, 1979; Brinton, 2001; Keenan and Schieffelin, 1976), whether verbal (e.g., address terms, locating directives or notice verbs such as *look* or *listen*, interrogatives, demonstratives, and imperatives) or non-verbal (e.g., pointing, touch, snapping gesture, (single) handclaps, the Raised Index Finger gesture, and throat clearing). One of the motivations for using attention-drawing devices may be a speaker's sense that they are not being listened to (e.g., Aijmer, 2010; Romero Trillo, 1997). This motivation was particularly evident in contexts where the addressee conveyed cues of disengagement (Excerpts 7 and 8), in which the addressee-directed index-finger pointing gesture was employed effectively as an attention-drawing device. In these cases, the pointer sought positive evidence of the addressee's attention in order to be reassured of their involvement in the interaction. Regaining attention was also relevant in the context of displaying the accomplishment of some cognitive process (Excerpts 9 and 10). In these cases, hesitations or disfluencies frequently arise, involving temporary suspension of flowing speech (e.g., Kosmala, 2021), thus potentially distracting the recipient's attention.

The addressee-directed pointing gestures share additional features with other gestural and verbal attention-getters. The study showed that, similar to particular attention-getters, the gesture in question can be employed as an abrupt method of interruption (Excerpt 11) or as an attempt to elicit a (minimal) response from the addressee (Excerpts 12 and 13). The participants used the gesture to enter or intrude into the discourse in order to convey the content that they had just recalled or to request clarification or confirmation. In a similar vein, the Raised Index Finger gesture (Inbar, 2022; *cf.* Uskokovic and Talehgani-Nikazm, 2022) and various notice verbs (e.g., *look* and *listen*; Aijmer, 2010; Keenan et al., 1987), in addition to drawing attention to the message, were used to take the floor or to interrupt (e.g., Aijmer, 2010; Brinton, 2001).

Another reason for the use of attention-getting devices has been attributed to the need to emphasize an important part of the utterance in order to ensure that the recipient understands the message correctly (Romero Trillo, 1997; see also Brinton, 2001; Keenan et al., 1987). Bavelas et al. (1995) noted that speakers gesture toward another person to emphasize part of their speech, while Yasui (2023) observed that speakers may produce a pointing gesture to focus on one particular part of their utterance. Some scholars have pointed out that the need to draw attention to an utterance could be motivated by disagreements and that verbal attention-drawing devices could be used in argumentative contexts (e.g., Aijmer, 2010; Brinton, 2001). In fact, in the majority of the cases, the index-finger pointing at the addressee occurred in contexts that involved performing an action that did not align with or support the co-participant's stance or perspective.

Moreover, this study has shown that disaffiliation was typically conveyed in tandem with establishing and sustaining the status and stance of epistemic authority. The association between epistemic authority and both verbal and non-verbal attention-drawing devices has been established in the literature. For example, Fairclough (2001) observed that speakers can assert their authority via the frequent use of the verbal attention-drawing device *look*. The connection between epistemic authority, opposition, and pointing has been highlighted in relation to the Raised Index Finger gesture (Inbar, 2022).

Another question to be addressed is what, nonetheless, is driving this direction of pointing. What motivates a speaker to point **at the addressee** in dyadic interactions, especially when the addressee's reference is not indicated in speech? It can be assumed that this direction of pointing can be prompted by an **appeal** shared by all contexts, in that a favorable response or consideration is being sought from the addressee, rather than merely providing information. This appeal can encompass requests for engagement, clarification, suspension, or acknowledgment of the speaker's perspective. The aim of engaging the addressee in the interaction brings non-referential pointing at the addressee closer to the category of summons (e.g., Pillet-Shore, 2018; Schegloff, 1968, 2002), designed to invite or prompt a response from a co-participant. However, the relation between the gesture and the category of summons warrants a more thorough examination.

6 Conclusion

This study focused on non-referential index-finger pointing at the addressee, which is a gesture that indexes the addressee for interactional purposes that extend beyond merely indicating a reference. The study revealed the contexts in which non-referential pointing at the addressee occurs in Hebrew talk-in-interaction. To elaborate on its pragmatic functions, the analysis was grounded in a detailed examination of the examples from each context, including the identification and analysis of the multimodal gestalts of which the gesture is part, as well as an exploration of the gesture's formation and semiosis. The study proposed that the gesture may primarily serve as a means of capturing attention, showing that it shares several characteristics with other verbal and non-verbal attention-drawing devices. Revealing these characteristics and their interrelationships makes it possible to better understand the phenomena of non-referential pointing, pointing directed at the addressee, and attention-drawing devices in general.

The recurrent use of the gesture in particular contexts may plausibly lead to expectations from the addressee concerning other persistent aspects of the information provided, action accomplished, or stance taken by the pointer in such contexts. Consequently, these meanings may become conventionalized and recognizable via this gesture. For example, being frequently produced in contexts where a speaker opposes a co-participant's stance—while also typically maintaining convergent status and a stance of epistemic authority—the gesture can be reanalyzed as a cue whereby conveying a negative or epistemic stance is recognized. By elaborating on such cases, this study contributes to the growing body of knowledge about multimodal stance-taking (e.g., Andries et al., 2023; Inbar, 2022; Inbar and Maschler, 2023; Newman et al., 2023; Shor and Marmorstein, 2022).

Furthermore, it has been indicated that both gestural and verbal deictic elements may evolve into stance markers via discourse deixis (e.g., Inbar, 2022; Shor and Inbar, 2019). For example, Heine and Kuteva (2002) note that, in various languages, demonstratives provide fertile ground for grammaticalization processes that can lead to the creation of various grammatical elements, including focus particles. These studies reinforce the conceptual link among deixis, focus, and stance. This connection is also highlighted herein.

The study showed that the addressee-directed pointing gesture may entail the blending of two or more semantic or functional categories to form a multifaceted hybrid sign. Moreover, the gesture can be accounted for from different perspectives, while various factors (e.g., morphological differences, seating arrangements, the number of participants) may have an impact on its employment. To move toward a more generalized analysis of the phenomenon and to develop distributional rules, the analysis should be expanded by considering new data. Finally, further research is needed to explore potential cultural differences, variations in genre and gender, and a more detailed examination of the similarities and differences among other verbal and non-verbal means used to attract attention in interaction.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The datasets presented in this article cannot be publicly available for privacy reasons. Requests to access these datasets should be directed to Anna Inbar, inbara9@gmail.com.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

AI: Conceptualization, Writing – original draft, Writing – review & editing, Methodology.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Israel Science Foundation, grants nos. 1233/16 and #941/20 to Yael Maschler. The publication of this article was supported by the Levinsky-Wingate Academic College.

Acknowledgments

I am grateful to the reviewers for their thorough feedback and valuable insights that have enhanced this manuscript.

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

Aijmer, K. (2010). "The attention-getting devices look, see and listen" in The Prague school and theories of structure (interfacing science, literature, and the humanities/ ACUME 2(1)). eds. M. Procházka, M. Malá and P. Šaldová (Göttingen: V&R unipress GmbH), 163–176.

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). Multimodal stance-taking in interaction––a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Atkinson, M. (1979). "Prerequisites for reference" in Developmental pragmatics. eds. E. Ochs and B. B. Schieffelin (London: Academic Press), 229–249.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15, 415–419. doi: 10.1111/j.0956-7976.2004.00694.x

Bardenstein, R., and Shor, L. (2019). Suspending progressivity – rega "moment" and ʃnija "second". Helkat lashon – a journal for theoretical and applied. *Linguistics* 52, 114–133.

Bavelas, J., and Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture* 17, 98–127. doi: 10.1075/gest.00012.bav

Bavelas, J. B., Chovil, N., Coates, L., and Roe, L. (1995). Gestures specialized for dialogue. *Personal. Soc. Psychol. Bull.* 21, 394–405. doi: 10.1177/0146167295214010

Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse Process.* 15, 469–489. doi: 10.1080/01638539209544823

Ben-Moshe, Y. M., and Maschler, Y. (Forthcoming). Tracing the semantic development of Hebrew clicks.

Ben-Moshe, Y. M., and Maschler, Y. (2024). Request for confirmation sequences in Hebrew: Open Linguistics. 10. doi: 10.1515/opli-2024-0028

Bowles, L. R. (2017). Doing the snap: storytelling and participatory photography with women porters in Ghana. *Vis. Anthropol. Rev.* 33, 107–118. doi: 10.1111/var.12129

Bressem, J., and Ladewig, S. H. (2011). Rethinking gesture phases: articulatory features of gestural movement? *Semiotica* 2011, 53–91. doi: 10.1515/semi.2011.022

Brinton, L. J. (2001). From matrix clause to pragmatic markers. The history of look forms. *J. Hist. Pragmat.* 2, 177–199. doi: 10.1075/jhp.2.2.02bri

Carranza, A. V. (2016). Remembering and noticing: a conversation-analytic study of 'ah' in Mexican Spanish talk. *Span. Context* 13, 212–236. doi: 10.1075/sic.13.2.03vaz

Chafe, W. (1994). Discourse, consciousness, and time. Chicago: University of Chicago Press.

Christidou, D. (2018). Art on the move: the role of joint attention in visitors' encounters with artworks. *Learn. Cult. Soc. Interact.* 19, 1–10. doi: 10.1016/j.lcsi.2018.03.008

Clark, H. H. (2003). "Pointing and placing" in Pointing: where language, culture, and cognition meet. ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum), 243–268.

Clark, H. H., and Brennan, S. A. (1991). "Grounding in communication" in Perspectives on socially shared cognition. eds. L. B. Resnick, J. M. Levine and S. D. Teasley (Washington: APA Books), 127–149.

Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *J. Verbal Learn. Verbal Behav.* 22, 245–258. doi: 10.1016/S0022-5371(83)90189-5

Cooperrider, K. (2011). Reference in action: Links between pointing and language. PhD thesis. San Diego: University of California.

Cooperrider, K. (2014). Body-directed gestures: pointing to the self and beyond. *J. Pragmat.* 71, 1–16. doi: 10.1016/j.pragma.2014.07.003

Cooperrider, K. (2016). The co-organization of demonstratives and pointing gestures. *Discourse Process.* 53, 632–656. doi: 10.1080/0163853X.2015.1094280

Cooperrider, K. (2023). Fifteen ways of looking at a pointing gesture. *Public J. Semiot.* 10, 40–84. doi: 10.37693/pjos.2023.10.25120

Cooperrider, K. (Forthcoming). The design of pointing. doi: 10.31234/osf.io/dba4c

Cooperrider, K., Slotta, J., and Núñez, R. (2018). The preference for pointing with the hand is not universal. *Cogn. Sci.* 42, 1375–1390. doi: 10.1111/cogs.12585

Cornish, F. (2011). 'Strict' anadeixis, discourse deixis and text structuring. *Lang. Sci.* 33, 753–767. doi: 10.1016/j.langsci.2011.01.001

Cornish, F. (2012). Micro-syntax, foregrounding and backgrounding in discourse – when indexicals target discursively subsidiary information. *Belgian J. Linguist.* 26, 6–34. doi: 10.1075/bjl.26.01cor

Couper-Kuhlen, E., and Selting, M. (2018). Interactional linguistics. Cambridge: Cambridge University Press.

Darwin, C. (1998) in The expression of the emotion in man and animals. ed. P. Ekman (London: Fontana Press).

Du Bois, J. W. (2012). Representing discourse. Unpublished manuscript, linguistics department, University of California at Santa Barbara (fall 2012 version). Available at: https://www.linguistics.ucsb.edu/research/santa-barbara-corpus

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., and Paolino, D. (1992). Discourse transcription: Santa Barbara papers in linguistics. California: University of California.

Eco, U. (1976). A theory of semiotics. Bloomington: Indiana University Press.

Enfield, N. J. (2009). The anatomy of meaning: Speech, gesture, and composite utterances. Cambridge: Cambridge University Press.

Enfield, N. J., Kita, S., and de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *J. Pragmat.* 39, 1722–1741. doi: 10.1016/j.pragma.2007.03.001

Fairclough, N. (2001). Language and power. Harlow: Longman.

Fauconnier, G., and Turner, M. (2002). The way we think: Conceptual blending and the Mind's hidden complexities. New York: Basic Books.

Goodwin, C. (2003). "Pointing as situated practice" in Pointing: Where language, culture, and cognition meet. ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum), 217–241.

Goodwin, C. (2018). Co-operative action. Cambridge: Cambridge University Press.

Goodwin, M. H., and Goodwin, C. (1986). Gesture and coparticipation in the activity of serching a word. *Semiotica* 62, 51–75.

Haviland, J. B. (2000). "Pointing, gesture spaces and mental maps" in Language and gesture. ed. D. McNeill (Cambridge: Cambridge University Press), 13–46.

Healy, C. (2012). Pointing to show agreement. *Semiotica* 2012, 175–196. doi: 10.1515/sem-2012-0073

Heine, B., and Kuteva, T. (2002). World lexicon of grammaticalization. Cambridge: Cambridge University Press.

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement" in Structures of social action. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Heritage, J. (2012). Epistemics in action: action formation and territories of knowledge. *Res. Lang. Soc. Interact.* 45, 1–29. doi: 10.1080/08351813.2012.646684

Heritage, J. (2016). On the diversity of 'changes of state' and their indices. *J. Pragmat.* 104, 207–210. doi: 10.1016/j.pragma.2016.09.007

Hindmarsh, J., and Heath, C. (2000). Embodied reference: a study of deixis in workplace interaction. *J. Pragmat.* 32, 1855–1878. doi: 10.1016/S0378-2166(99)00122-8

Hirst, W., and Echterhoff, G. (2012). Remembering in conversations: the social sharing and reshaping of memories (January 2012). *Annu. Rev. Psychol.* 63, 55–79. doi: 10.1146/annurev-psych-120710-100340

Holler, J. (2010). "Speaker's use of interactional gestures as markers of common ground" in Gesture in embodied communication and human-computer interaction. eds. S. Kopp and I. Wachsmuth (Berlin/Heidelberg: Springer), 11–22.

Hopper, P. J. (2021). "You turn your back and there's somebody moving in" syntactic anacrusis in spoken English. *Interact. Linguist.* 1, 64–89. doi: 10.1075/il.20009.hop

Inbar, A. (2020). On the functions of lists in discourse: the analysis of gestures coordinated with list constructions in spoken Israeli Hebrew. *Proc. Annu. Meet. Haiim Rosén Israeli Linguist. Soc.* 22, 69–84.

Inbar, A. (2022). The raised index finger gesture in Hebrew multimodal interaction. *Gesture* 21, 264–295. doi: 10.1075/gest.21001.inb

Inbar, A. (Forthcoming). Visualizing discourse functions: the analysis of gestures associated with List Constructions in Israeli Hebrew.

Inbar, A., and Maschler, Y. (2023). Shared knowledge as an account for disaffiliative moves: Hebrew ki 'because'-clauses accompanied by the palm up open hand gesture. *Res. Lang. Soc. Interact.* 56, 141–164. doi: 10.1080/08351813.2023.2205302

Ishino, M. (2009). Roles of gestures pointing to the addressee in Japanese face-to-face interaction: attaining cohesion via metonymy. *Japanese Korean Linguist.* 16, 264–278.

Jefferson, G. (1987). "Sequential aspects of story telling in conversation" in Studies in the organization of conversational interaction. ed. J. Schenkein (New York: Academic Press), 213–248.

Keenan, E., and Schieffelin, B. B. (1976). "Topic as a discourse notion: a study of topic in the conversations of children and adults" in Subject and topic. ed. C. N. Li (New York: Academic Press, Inc.), 335–384.

Keenan, E. O., Schieffelin, B. B., and Platt, M. (1987). "Questions of immediate concern" in Questions and politeness: strategies in social interaction. ed. E. N. Goody (Cambridge: Cambridge University Press), 44–55.

Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance" in Nonverbal communication and language. ed. M. R. Key (Berlin: Mouton), 207–277.

Kendon, A. (2004). Gesture: visual action as utterance. Cambridge: Cambridge University Press.

Kita, S. (2003). "Pointing: a foundational building block of human communication" in Pointing: where language, culture, and cognition meet. ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum), 1–8.

Kita, S., van Gijn, I., and van der Hulst, H. (1998). "Movement phases in signs and cospeech gestures and their transcription by human encoders" in Gesture, and sign language in human-computer interaction. eds. I. Wachsmuth and M. Fröhlich (Berlin/Heidelberg: Springer), 23–35.

Kosmala, L. (2021). A multimodal contrastive study of (dis)fluency across languages and settings: towards a multidimensional scale of inter-(dis)fluency. PhD thesis,. Paris III: Université de la Sorbonne nouvelle.

Lakoff, G. (1987). Women, fire and dangerous things: what categories reveal about the mind. Chicago: University of Chicago Press.

Langacker, R. W. (2008). Cognitive grammar: a basic introduction. Oxford: Oxford University Press.

Langton, S. R. H., and Bruce, V. (2000). You must see the point: automatic processing of cues to the direction of social attention. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 747–757. doi: 10.1037//0096-1523.26.2.747

Levinson, S. C. (2006). "On the human 'interaction engine'" in Roots of human sociality: Culture, cognition and interaction. eds. N. J. Enfield and S. C. Levinson (Berg: Oxford), 39–69.

Maschler, Y. (2017). "Blurring the boundaries between discourse marker, pragmatic marker, and modal particle: the emergence of Hebrew loydea/loydat ("I dunno MASC/FEM") from interaction" in Pragmatic markers, discourse markers and modal particles: new perspectives. eds. A. Sansò and C. Fedriani (Amsterdam: John Benjamins), 37–69.

Maschler, Y., Polak-Yitzhaki, H., Aghion, G., Fofliger, O., Wildner, N., Ben-Moshe, Y. M., et al. (2024). The Haifa multimodal Corpus of spoken Hebrew.

McNeill, D. (1992). Hand and mind. Chicago: University of Chicago Press.

McNeill, D., Cassell, J., and Levy, E. T. (1993). Abstract deixis. *Semiotica* 95, 5–9. doi: 10.1515/semi.1993.95.1-2.5

Middleton, D., and Edwards, D. (1990). "Conversational remembering: a social psychological approach" in Collective remembering. eds. D. Middleton and D. Edwards (London: SAGE), 23–45.

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Stud.* 9, 194–225. doi: 10.1177/1461445607075346

Mondada, L. (2012). "Deixis: an integrated interactional multimodal analysis" in Prosody and embodiment in interactional grammar. eds. P. Bergman and J. Brenning (Berlin: De Gruyter), 173–206.

Mondada, L. (2014). "Pointing, talk, and the bodies: reference and joint attention as embodied interactional achievements" in From gesture in conversation to visible action as utterance: essays in honor of Adam Kendon. eds. M. Seyfeddinipur and M. Gullberg (Amsterdam/Philadelphia: John Benjamins Publishing Company), 95–124.

Mondada, L. (2016). Challenges of multimodality: language and body in social interaction. *J. Socioling.* 20, 336–366. doi: 10.1111/josl.1_12177

Mondada, L. (2019). Conventions for multimodal transcription. Available at: https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/home/Personen/Mondada/Unterordner/Mondada_conv_multimodality.pdf (Accessed June 30, 2024).

Newman, G., Inbar, A., and Shor, L. (2023). Cutting off inappropriate formulations: a disalignment practice in Hebrew face-to-face interaction. 18th International Pragmatics Conference (IPrA), Université Libre de Bruxelles (ULB), Brussels.

Pillet-Shore, D. (2018). How to begin. *Res. Lang. Soc. Interact.* 51, 213–231. doi: 10.1080/08351813.2018.1485224

Pomerantz, A. (1986). Extreme case formulations: a way of legitimizing claims. *Hum. Stud.* 9, 219–229. doi: 10.1007/BF00148128

Reisenzein, R., Meyer, W. U., and Niepel, M. (2012). "Surprise" in Encyclopedia of human behavior. ed. V. S. Ramachandran. *2nd* ed (New York: Academic Press), 564–570.

Romero Trillo, J. (1997). Your attention, please: pragmatic mechanisms to obtain the addressee's attention. *J. Pragmat.* 28, 2005–2221.

Sacks, H. (1992). Lectures on conversation. Oxford: Basil Blackwell.

Sacks, H., and Schegloff, E. A. (1979). "Two preferences in the organization of reference to persons in conversation and their interaction" in Everyday language: studies in ethnomethodology. ed. G. Psathas (New York: Irvington), 15–21.

Schegloff, E. A. (1968). Sequencing in conversational Openings1. *Am. Anthropol.* 70, 1075–1095. doi: 10.1525/aa.1968.70.6.02a00030

Schegloff, E. A. (2002). "Opening sequencing" in Perpetual contact: mobile communication, private talk, public performance. eds. J. E. Katz and M. A. Aakhus (Cambridge: Cambridge University Press), 325–385.

Scott, S. (2002). Linguistic feature variation within disagreements: an empirical investigation. *Text* 22, 301–328. doi: 10.1515/text.2002.011

Shor, L., and Inbar, A. (2019). The meaning of zehu in spoken Israeli Hebrew: a corpus based analysis of its interjectional function. *Scand. Lang. Stud.* 10, 131–151. doi: 10.7146/sss.v10i1.114675

Shor, L., Inbar, A., and Izre'el, S. (Forthcoming). From direct evidentiality to discursive corroborative evidence and beyond: the case of the particle hine in spoken Hebrew.

Shor, L., Inbar, A., and Izre'el, S. (Forthcoming). Revisiting presentative particles: multimodal study of the Hebrew particle hine.

Shor, L., and Marmorstein, M. (2022). The embodied modification of formulations: the Guating gesture (QG) in Israeli-Hebrew discourse. *J. Pragmat.* 192, 22–40. doi: 10.1016/j.pragma.2022.01.019

Sidnell, J. (2007). 'Look'-prefaced turns in first and second position: launching, interceding and redirecting action. *Discourse Stud.* 9, 387–408. doi: 10.1177/1461445607076204

Streeck, J. (2017). Self-making man: a day of action, life, and language. Cambridge: Cambridge University Press.

Stukenbrock, A. (2014). Pointing to an 'empty' space: Deixis am Phantasma in face-to-face interaction. *J. Pragmat.* 74, 70–93. doi: 10.1016/j.pragma.2014.08.001

Uskokovic, B., and Talehgani-Nikazm, C. (2022). Talk and embodied conduct in word searches in video-mediated interactions. *Soc. Interact. Video Based Stud. Hum. Soc.* 5. doi: 10.7146/si.v5i2.130876

Wilkins, D. (2003). "Why pointing with the index finger is not a universal (in sociocultural and semiotic terms)" in Pointing: where language, culture, and cognition meet. ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum), 171–216.

Will, I. (2021). Recurrent gestures of Hausa speakers. Leiden: Brill.

Yasui, E. (2017). Chokuzen no hanashite o yubisasu koto: Chokuzen no hatsuwa to no kanren o shimesu shigen toshite no yubisashi [Pointing at a previous speaker: pointing gestures as a resource to make a direct link with a previous utterance]. *Jpn. J. Lang. Soc.* 20, 131–145. doi: 10.19024/jajls.20.1_131

Yasui, E. (2023). Sequence-initial pointing: spotlighting what just happened as a cause of a new sequence. *Discourse Stud.* 25, 409–429. doi: 10.1177/14614456221132464

Ziv, Y. (2016). "Selected comment expressions from CoSIH – distribution, functions and analytical challenges" in Studies in Spoken Hebrew. ed. E. Gonen (Tel Aviv: Tel Aviv University Press), 525–550.

Appendix

Transcription conventions

(Following Chafe, 1994; Du Bois et al., 1992; Du Bois, 2012 and adapted for Hebrew (Maschler, 2017))

.. – perceptible pause of less than 0.1 s

... – average pause ($0.1 \leq x < 1.0$ s)

.... – pause ($1.0 \leq x < 1.5$ s)

.... – pause ($1.5 \leq x < 2.0$ s).

(3.56) – measured pause of 3.56 s.

, – comma at end of line –mid-level, mid-rise, mid-fall intonation, regularly understood in Hebrew as 'more to come'

. – period at end of line – low fall intonation, regularly understood in Hebrew as final.

? – question mark at end of line – high rising intonation, regularly understood in Hebrew as.

final and seeking response from interlocutor.

?, – question mark followed by comma – rising intonation, regularly understood in Hebrew as projecting 'more to come' while seeking response from interlocutor.

! – exclamation mark at end of line – final exclamatory intonation.

ø – lack of punctuation at end of line – a fragmentary intonation unit, one which never reached completion.

-- two hyphens – elongation of preceding sound.

underlined syllable – primary stress of intonation unit.

boldfaced syllable – secondary stress of intonation unit.

@ – a burst of laughter (each additional @ symbol denotes an additional burst).

square bracket to the left of two consecutive lines indicates.

beginning of overlapping speech, two speakers talking at once.

inverted bracket + alignment such that the right of the top line.

is placed over the left of the.

bottom line indicates latching, no interturn pause.

Musical notation as necessary: e.g., *acc* – *accelerando* (progressively faster).

(in regular brackets) – nonverbal action constituting a turn.

{in curly brackets} – transcriber's comments.

'– uninverted quotation mark indicates the glottal stop phoneme.

- one hyphen – bound-morpheme boundary.

/words within slashes/ indicate uncertain transcription.

Transcription of embodied conduct.

(following Mondada, 2019)

Descriptions of embodied conduct are delimited in between symbols.

± ± Two identical symbols (one symbol per participant and per type of conduct) are synchronized with corresponding stretches of talk.

#----> Described embodied conduct continues across subsequent lines

---- > # until the same symbol is reached.

..... Action's preparation.

--------- Action's apex is reached and maintained.

#----- > l.12 Described embodied conduct continues until line 12 of transcript.

#----->> Described embodied conduct continues beyond end of excerpt.

> > –---- Described embodied conduct begins before the excerpt's beginning.

* Exact position in the utterance in which a video caption was made.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 27 January 2025
DOI 10.3389/fpsyg.2025.1499538

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Wenxing Yang,
Yangzhou University, China
Lin Wang,
Capital Institute of Pediatrics, China
Shu-Yu Huang,
Defense Language Institute Foreign Language Center, United States

*CORRESPONDENCE
Chao Sun
chaosun@pku.edu.cn

RECEIVED 21 September 2024
ACCEPTED 14 January 2025
PUBLISHED 27 January 2025

CITATION
Zheng X and Sun C (2025) Understanding perspective-taking in multiparty conversations: insights from Mandarin nouns. *Front. Psychol.* 16:1499538. doi: 10.3389/fpsyg.2025.1499538

COPYRIGHT
© 2025 Zheng and Sun. 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.

Understanding perspective-taking in multiparty conversations: insights from Mandarin nouns

Xiaobei Zheng[1] and Chao Sun[2]*

[1]College of International Studies, Shenzhen University, Shenzhen, China, [2]School of Chinese as a Second Language, Peking University, Beijing, China

Individuals frequently adopt others' perspectives both when interpreting language and when formulating their own responses in conversation. This experiment tested how participants used perspective information to resolve references for bare nouns in Mandarin. Specifically, it explored whether, when faced with two interlocutors, participants distinguished between each individual's perspective or considered both as a whole. Using a classical referential game, the study manipulated the visual perspectives of two partners. In Experiment 1, both speakers had the same seating direction and visual field, and the results showed that participants equally took their perspectives into account above chance levels, providing a baseline finding for referential resolution of Mandarin bare nouns in perspective-taking studies. In Experiment 2, both speakers had the same seating direction but one of them shared the larger portion of visual field with the participants. The results showed that participants took the perspectives of the two speakers independently, while also comparing the perspectives of both interlocutors to facilitate quicker and more accurate referential resolution. These findings demonstrate that perspective-taking is a complex and dynamic process, providing evidence for the study of perspective-taking in Mandarin and contributing insights into comprehension processing in multiparty conversations.

KEYWORDS

perspective-taking, common ground, Mandarin, bare noun, referential resolution

1 Introduction

1.1 Common ground in communication

During daily communication, in addition to employing linguistic knowledge to construct discourse, interlocutors consider the physical background, the information known to other parties, and other contextual factors. One fundamental factor is the common ground shared between communication participants (Stalnaker, 1978). Common ground refers to the knowledge, beliefs, and information mutually understood by interlocutors. In language philosophy, it includes a deeper dimension, such as the awareness interlocutors have of their partners' awareness of this shared knowledge, beliefs, and information (Schiffer, 1972; Clark and Marshall, 1978; Aumann, 1976). In psycholinguistics, researchers mainly investigate how shared information affects online language production and comprehension, a process referred to as perspective-taking during conversation.

Psycholinguistic studies often explore how individuals use both privileged and common ground information during referential processing. Privileged ground refers to knowledge known only to one interlocutor, while common ground consists of information shared by both. Research shows that although common ground facilitates faster and more accurate comprehension, participants occasionally consider their

privileged ground (Heller and Brown-Schmidt, 2023; Heller et al., 2016; Mozuraitis et al., 2018; Ryskin et al., 2020; Keysar et al., 2000; Chambers et al., 2008; Epley et al., 2004), particularly when interpreting questions posed by others (Brown-Schmidt et al., 2008).

1.2 Strategies in perspective-taking

To resolve the debate regarding the use of common vs. privileged ground in conversation, psycholinguistic studies have focused on the timing of processing. Some studies propose an initial egocentric strategy, where people prioritize their own knowledge and attentional states and then adjust based on common ground (Keysar et al., 2000; Keysar et al., 1998). Other research suggests that individuals begin with an initial consideration of common ground, treating it as one factor in top-down language processing that operates alongside other linguistic factors (Hanna et al., 2003; Heller et al., 2012; Nadig and Sedivy, 2002; Sikos et al., 2019; Rubin et al., 2011).

Evidence supporting both proposals has accumulated, but results are often influenced by experimental design. Consequently, some theories propose that perspective-taking does not follow an "all-or-none" approach. Certain models blur the distinction between privileged and common ground. For example, the memory-based model suggests that ground-related information is stored as a general episodic memory, which is later retrieved through resonance with relevant cues during language processing (Horton and Gerrig, 2005a, 2005b, 2016). If, for instance, Tom received a phone call from his boss while Mary was present, Mary would serve as a cue for Tom to retrieve the memory of the call later, even if the content of the call did not relate to their common ground. Similarly, when information lacks strong cues, individuals may rely on an egocentric strategy, even if that information exists in the common ground. This proposal helps resolve previous debates on the use of egocentric strategies versus common ground.

Another proposal, partner-specific processing, suggests that individuals establish different common grounds with different individuals. This proposal is also integrated with the memory-based model by emphasizing the unique role of humans. That is, among all kinds of memory cues, humans—as social agents—are the most powerful triggers, making information jointly attended to by both partners more easily retrievable for later use (Brown-Schmidt, 2012; Brown-Schmidt et al., 2015; Brown-Schmidt and Hanna, 2011). In the phone call example, if Tom had been watching TV when he answered the call, the TV show might act as a cue for later retrieval. However, compared to the TV show, Mary would be a stronger cue when recalling the content of the call.

Research on conceptual pacts supports this partner-specific explanation. Specifically, individuals tend to maintain the same expressions when conversing with the same partner but abandon them when interacting with a new partner (Metzing and Brennan, 2003; Brown-Schmidt, 2009; Kronmüller and Barr, 2007). Similar proposals have also been made in conversational theories. For instance, Pickering and Garrod (2004, 2013, 2021) emphasized cognitive alignment, where participants unconsciously align their language at various levels, such as syntax and vocabulary, to facilitate communication. The process of cognitive alignment also involves memory, integrating previous linguistic information with the partner's identity.

Philosophical discussions of common ground also address the influence of interlocutors-related information and the processing mechanisms associated with it. From a philosophical perspective, common ground entails an infinite regress, where each participant infers that their partner knows that they know a given proposition, continuing indefinitely. Moreover, common ground is seen as a dynamic entity that evolves throughout communication. Therefore, such processing places considerable cognitive demands on interlocutors. To address this challenge, Clark and Marshall introduced the concept of the co-presence heuristic, where common ground is established based on a shared physical environment, prior discourse, and cultural background (Clark, 1996; Clark and Marshall, 1981). Using these heuristics, speakers do not need to assess whether every piece of information is in or out of the common ground. For example, when discussing the only visible book between interlocutors, one might use minimal description, such as "the book" or even just "it." The physical co-presence narrows down the referential domain without requiring deeper assessment. This explanation parallels the concept of joint attention in developmental psychology, which describes how infants and adults coordinate their focus on the same object, facilitating accurate reference resolution (Tomasello and Farrar, 1986). In both cases, physical co-presence contributes to establishing common ground between interlocutors.

These heuristics also take cultural and social background factors into account. For instance, when a doctor explains an illness to a patient, they may avoid using medical jargon, but might use more technical language when speaking to a colleague. This social and cultural background can be seen as a holistic strategy in which individuals adopt others' perspectives based on their social identity or stance, allowing for quicker evaluations. The assessment of overall similarity facilitates communication by promoting mutual understanding and reducing misinterpretations (Runkel, 1956). These philosophical and empirical discussions show that people integrate information about the communicative partner's identity, social background, and the surrounding context to form certain expectations and strategies that help them communicate more effectively.

1.3 Research questions in partner-specific processing

In summary, the aforementioned explanations suggest a dynamic, non-binary approach to processing perspective in language comprehension. Whether focusing on partner-specific or cultural heuristics, these studies emphasize the importance of interlocutor-related information. However, the interlocutor-related information is quite broad.

First, interlocutor-related information has different categories at the social and cognitive levels. In cultural heuristics, the partner is seen as part of a social group with a certain background. For example, the conversational partner could be a patient of the doctor or a colleague of the doctor. However, in memory-based and cognitive alignment models, the information about the interlocutor shifts to memory or perceptual priming, which is an automatic process. For instance, when conversation partners sit face-to-face, factors such as the direction of their seats or their line of sight might involve only the automatic processing of the physical environment.

Second, how the partner-related information or similarity between the partner and oneself influences perspective-taking remains unclear. It may lead to either an egocentric or collective strategy. Studies suggest that people overestimate the mental states of familiar individuals and display more egocentrism towards the perspectives of friends compared to strangers (Savitsky et al., 2011; Robbins and Krueger, 2005). The similarity at the cognitive level has also been examined experimentally. High information overlap between communicators can reduce communication effectiveness, as speakers tend to overestimate listeners' knowledge (Wu and Keysar, 2007a). As shared information increases, people are more likely to overestimate common ground (Zheng and Breheny, 2021). Research on child development provides similar evidence (Moll and Meltzoff, 2011; Moll et al., 2007, 2011). In these studies, two adult speakers engage with a child, each establishing distinct common ground through joint attention. The findings suggest that children under the age of two can differentiate between the common ground they share with different partners. However, in some cases, even when adults do not share information with children, children may still treat the adult as a knowledgeable party (Moll et al., 2011).

Last but not least, this study focuses on two additional aspects. First, it expands the scope of conversational analysis by moving beyond traditional dyadic interactions to triadic conversations. This shift introduces more complex scenarios, as the knowledge systems of the three participants involve overlapping layers of common ground between each pair. Second, the study emphasizes the comprehension processes of the addressee, rather than the production processes of the speaker. Specifically, it examines how the addressee interprets discourse based on the perspectives of others, as opposed to how speakers plan their discourse from their audience's perspective.

Based on this framework, we address two key research questions: (1) Does the addressee distinguish between the perspectives of the two speakers, leading to distinct understandings of their utterances? (2) Does the common ground shared between the two speakers and that between the addressee and each speaker interact in ways that shape the addressee's comprehension strategies? We hypothesize that addressees will adopt the perspectives of each speaker based on their respective visual fields; however, increased similarity between speakers and addressees may either enhance interpretive accuracy or lead to overly egocentric understandings. This study aims to provide empirical evidence to clarify these possibilities.

1.4 The present study

The present experiment employed a classic referential game in which speakers instructed participants to manipulate target objects on a shelf (Keysar et al., 2000). Participants and speakers sat on opposite sides of the shelf, which contained both transparent and opaque compartments, resulting in different sets of object information visible to the participants and the speakers. This setup allowed us to explore how participants adjust their interpretation of the speakers' instructions based on the varying levels of shared information.

Each participant engaged in the game with two confederate speakers, who alternated giving instructions. The purpose of this was to compare how participants analyse and integrate each speaker's perspective in the same communicative context. On the one hand, both confederates acted as speakers, seated opposite the participant, identical in their role and physical positioning, but distinct from the participant. This setup ensures that both speakers in the task have exactly the same role, i.e., as directors, and are seated in the same position, directly facing the participants. On the other hand, the two confederates may or may not share the same information with the participant. This setup allows for the manipulation of the common ground between the participants and each speaker. The study aims to investigate how the different levels of common ground influence participants' perspective-taking and language comprehension. We predict that participants will interpret different speakers' discourse based on the specific shared knowledge associated with each speaker, consistent with findings from previous studies on perspective-taking. Moreover, participants are also expected to partially integrate the perspectives of the two speakers because they share a common role in the discourse.

Additionally, the instructions were delivered in Mandarin, where nouns can stand alone without a determiner, a construction known as bare nouns. For example, the phrase "pick up dog" in Mandarin does not require a determiner. Bare nouns in Mandarin can convey definiteness, indefiniteness, or generality depending on the context (Li and Thompson, 1975, 1989). In previous referential studies conducted in English, the inclusion of the determiner "the" before a noun, as in "pick up the dog," typically leads the addressee to interpret the noun as definite, thereby encouraging perspective-taking. However, in Mandarin, bare nouns do not inherently imply such a preference. As a result, "pick up dog" can be interpreted as referring to a specific, indefinite, or generic dog. When multiple dogs are present, participants may reasonably consider picking up any, some, or all of them. This lack of grammatical constraint provides an opportunity to observe how participants make their choices in a more flexible linguistic environment.

In Experiment 1, since research on perspective-taking in Mandarin is relatively limited, we first conducted an exploratory experiment in which both confederates had identical identities and shared knowledge. We expected that participants will show no difference in their interpretation of the two confederates' instructions. Additionally, Experiment 1 will provide a baseline for participants' level of perspective-taking in this experimental design.

2 Experiment 1

2.1 Participants

Twenty-seven university students (14 female and 13 male, aged 17 to 24 years) participated in the experiment. Two female research assistants who were trained as confederates also took part in the experiment. All participants were recruited through campus advertisements and were informed that they would receive a cash reward regardless of their performance. They were also told they had the right to withdraw from the experiment at any time and still receive the reward. All participants are native Mandarin speakers and have lived in Mandarin-speaking communities since childhood. They were university students with similar levels of proficiency in Mandarin. None of the participants reported any cognitive, hearing, vision, or other relevant impairments.

2.2 Methods

2.2.1 Stimuli and design

The experiment was conducted in a quiet laboratory with a table placed in the center of the room. A shelf was positioned on the table. During the experiment, two cameras were used to record the participants' behavior and eye movements. The camera recording behavior was placed behind the laboratory curtains, while the camera recording eye movement trajectories was set up behind the shelf. In the experiment, participants sat at the front of the shelf, and two confederates were seated on the opposite side of the shelf. The shelf contained four compartments. Three compartments were transparent, making the objects in these compartments visible to both the participants and the two confederate speakers. These objects were shared objects. One compartment, however, was blocked by a partition, so the objects inside were visible only to the participants seated at the front. This object was referred to as an unshared object (as illustrated in Figure 1).

During the experiment, in each trial, the participants saw pictures of four objects placed in the four compartments. These objects included items such as animals, fruits, and vehicles. In each trial, the objects belonged to the same category; for instance, all animals. The arrangement of the cards in the compartments was manipulated according to competitor-type conditions. In the non-competitor condition, the four objects were different; for instance, a cat, a fish, a rabbit and a dog (Figure 1 left). In the competitor condition, the blocked compartment contained an object identical to one in a transparent compartment; for instance, a cat, a fish and two dogs (Figure 1 right). For comparative analysis, the non-competitor and competitor conditions were paired, meaning that the three shared objects in the transparent compartments and their positions were identical across conditions, with the only difference being whether the unshared object in the blocked compartment was a duplicate.

In each trial, the confederates' instructions were to "please pick… (the object name)." In the competitor condition, the critical object referred to the one present in both the blocked and transparent compartments. For participants seated in front of the shelf, this instruction was semantically ambiguous because it could refer to either the shared or the unshared object. In the non-competitor condition, the confederates used the same name, referring to the shared object in the transparent compartment without ambiguity. For example, in the sample display in Figure 1, the critical noun "dog" in the instruction "please pick (the) dog" could refer to either the shared or unshared dog in the competitor condition, but there was only one dog, which is shared, in the non-competitor condition. The participant and the two confederates could see each other, but to avoid the confederates' attention to objects giving hints to the participant, the confederates avoided making eye contact with the participant during the critical conversation.

To prevent participants from developing strategies during the experiment, such as immediately selecting the object in the transparent compartment when there was an identical object competitor in the blocked compartment, a filler condition was introduced. In this condition, the filler trials randomly used the same objects and arrangement as the experimental condition, but with two identical objects included. Unlike in the competitor condition, the two identical objects were not necessarily placed in the blocked compartment. Importantly, the confederates' instructions did not refer to these duplicate objects.

Moreover, the identity of the speakers constituted an additional experimental condition. Half of the trials were conducted under the instruction of Confederate 1 (C1), while the remaining half were under the guidance of Confederate 2 (C2), with both individuals providing identical instructions. These instructions were presented in a randomized sequence, with adjustments made for cases where two identical instructions from different speakers appeared within a span of three consecutive trials.

In summary, each participant received 48 trials. The competitor types served as a within-participant variable, with participants receiving 16 non-competitor trials and 16-competitor trials. An additional 16 filler trials were included. The confederate identity served as another within-participant variable, with each confederate providing 24 instructions.

2.2.2 Procedure

Upon arrival at the laboratory, participants were informed that they would engage in a cooperative task. They would

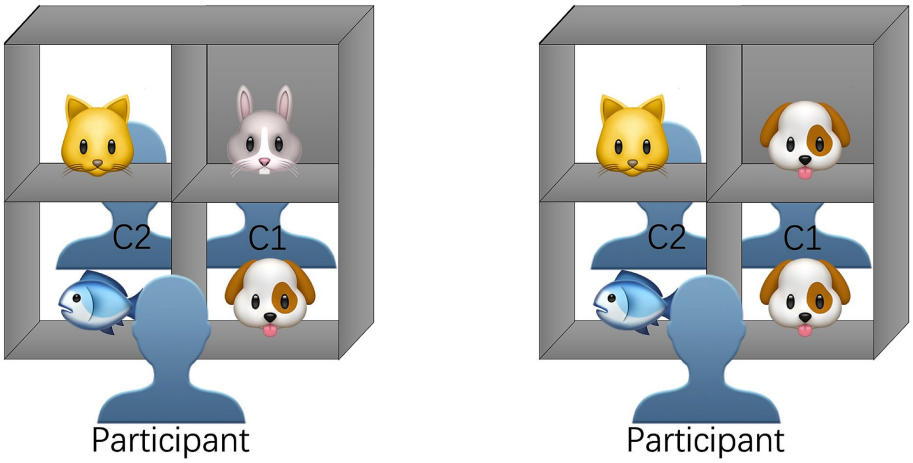


FIGURE 1
The display of the shelf and the seating arrangement of participants and confederates in Experiment 1.

be assigned as "operators," who were instructed to examine objects placed on a shelf by two other "instructors" (Confederate 1 and 2). The participants were told that the two confederates had previously functioned as operators and were now tasked with acting as instructors in the present task. To ensure the credibility of their conversations during the repeated picking task, the participants were also told that the two confederates were performing an additional task requiring the participants' assistance.

Next, to help participants familiarize themselves with the confederate's view, they were asked to sit on the opposite side of the shelf and complete several practice trials as instructors. Additionally, when the confederate acted as the operator, regardless of how many objects were visible to her, she would select only one object. This design aimed to convince participants that the confederates' extra task required just one object. In the formal test, no participant chose two or more objects.

Once the formal experiment began, confederates alternated in providing instructions. During the initial trials, they might feign hesitation, pretending that they were considering their own extra task. To minimize potential delays in participants' eye movements and behavioral data resulting from unfamiliarity during the initial phase, the first five trials of the experiment were all filler trials. This enabled participants to acquaint themselves with the task before engaging with the critical trials.

2.2.3 Coding

For the behavioral data, the experimenter recorded the objects chosen by the participants during the experiment. After the experiment, another research assistant re-coded the participant's choices based on the video recording. Finally, a third research assistant verified the records from the two coders and resolved any inconsistency using the video recordings. The participants' selection of target items was calculated. "Target items" refer to the object mentioned by the speaker in the non-competitor condition, as well as the object in the transparent grid mentioned by the speaker in the competitor condition. Each participant completed 8 trials for each condition, so the target choice proportion for each participant was calculated by dividing the number of target choices by 8, yielding four percentage scores for each participant under the four combinations of speaker and competitor types.

For the eye movement data, two trained research assistants coded the eye movement data based on the video recordings, with the coders blind to each trial's condition and the specific object arrangement. The coders performed frame-by-frame coding of the participant's gaze direction. The eye movement data were recorded at a frequency of 25 frames per second, which corresponded to fixation intervals of 40ms. Eye-tracking data concerning the target object were analyzed across the interval corresponding to the duration of the critical noun (e.g., "dog"), specifically from 200ms to 1200ms after word onset. The time window began at the 200ms mark following the onset of the critical word, in line with the standard assumption regarding the time necessary to program and execute an eye movement (Hallett, 1986). The time window ended at 1200ms due to the average latency at which participants began to select the objects. The coding followed the same method as Huang et al. (2013), with the fixation quadrant indicating the category of the object, which was further divided into target and non-target objects. The average proportion of looks to the target in each time bin was calculated by dividing target fixations by all fixations across the four objects.

2.3 Results

2.3.1 Behavioral data analyses

First, Wilcoxon signed ranks tests were conducted to examine the effect of competitor types on target choice, with Confederate 1 and 2 trials analysed separately. Therefore, competitor type (competitor vs. non-competitor) served as the independent variable, and the proportion of target selection served as the dependent variable (Figure 2).

In C1 (Confederate 1) trials, results showed a significant difference between competitor and non-competitor trials, $Z = -3.422$, $p = 0.001$, effect size = 0.88. Participants selected the target item more frequently in non-competitor trials (M = 1, SD = 0) compared to competitor trials (M = 0.78, SD = 0.25). Similarly, in C2 (Confederate 2) trials, a significant effect of competitor type was also observed, $Z = -3.179$, $p = 0.001$, effect size = 0.75. As in condition C1, participants were more likely to select the target item in non-competitor trials (M = 1, SD = 0) than in competitor trials (M = 0.76, SD = 0.32). These findings indicate that the presence of a competitor item reduced the likelihood of participants selecting the target item.

Secondly, the effects of speaker identity in competitor trials was explored by a Wilcoxon signed ranks test. The results showed that no significant difference between the target choice proportions in C1and C2 during competitor trials, $Z = -0.752$, $p = 0.452$, effect size = 0.069. This result indicates that the presence of a competitor similarly impacted participants' target selection across both speakers, with no observable variation between the two.

Thirdly, participants' target choice in the competitor condition was compared to the chance level of 0.5. Two separate one sample Wilcoxon signed ranks tests were performed for C1 and C2 trials. Results showed that for both C1 and C2, participants' target choice was significantly higher than chance, C1: $Z = 3.893$, $p < 0.001$, effect size = 1.12; C2: $Z = 3.527$, $p < 0.001$, effect size = 0.813, suggesting that participants were able to reliably select the target item despite the presence of a competitor item, significantly exceeding chance-level performance.

The results suggest that participants utilize perspective information to modulate the referential domain of target nouns to some extent. The word labels for targets were bare nouns. In Mandarin, bare nouns can refer to definite, indefinite or generic candidates, which implies a broader referential domain. However, participants' referential resolution to the shared items is above chance, showing that they used perspective cues to solve the referential resolution.

2.3.2 Eye-tracking data analyses

A general linear mixed-effects model was conducted using the lmer function in R to examine the impact of competitor type and speaker on participants' eye movements. The model included competitor type (competitor vs. non-competitor) and speaker (C1 vs. C2) as fixed effects, while participants and items were entered as random effects. The dependent variable was the proportion of target fixation. The fixation towards the target was graphed over time in Figure 3.

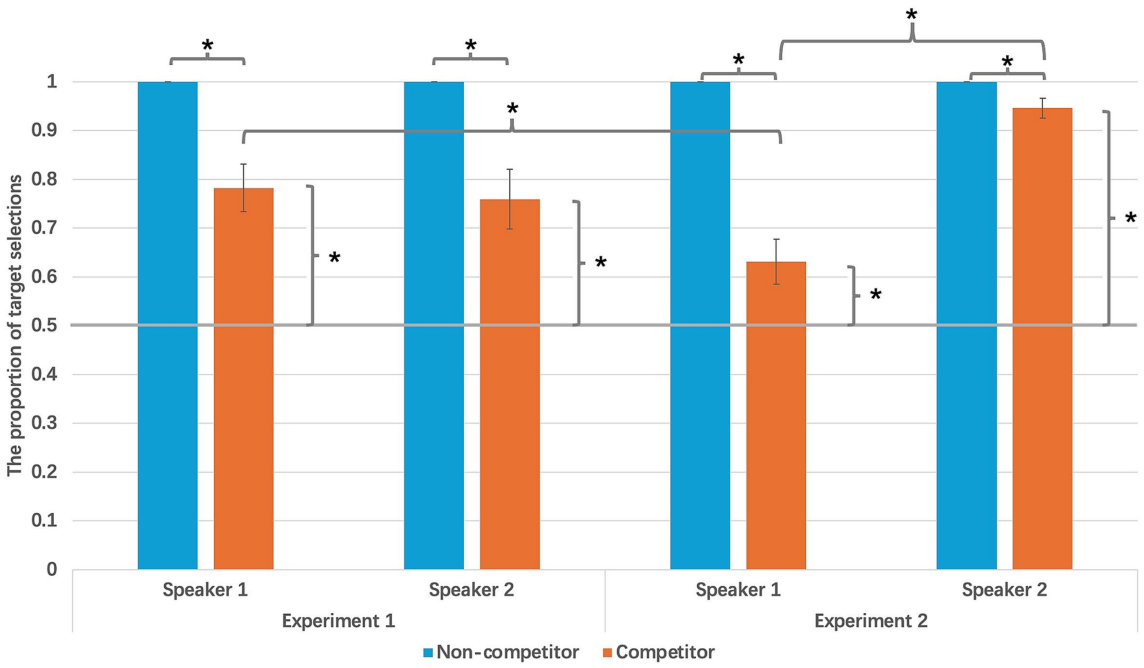


FIGURE 2
Proportion of participants' selection of the target object from two confederates across two competitor types in Experiments 1 and 2.

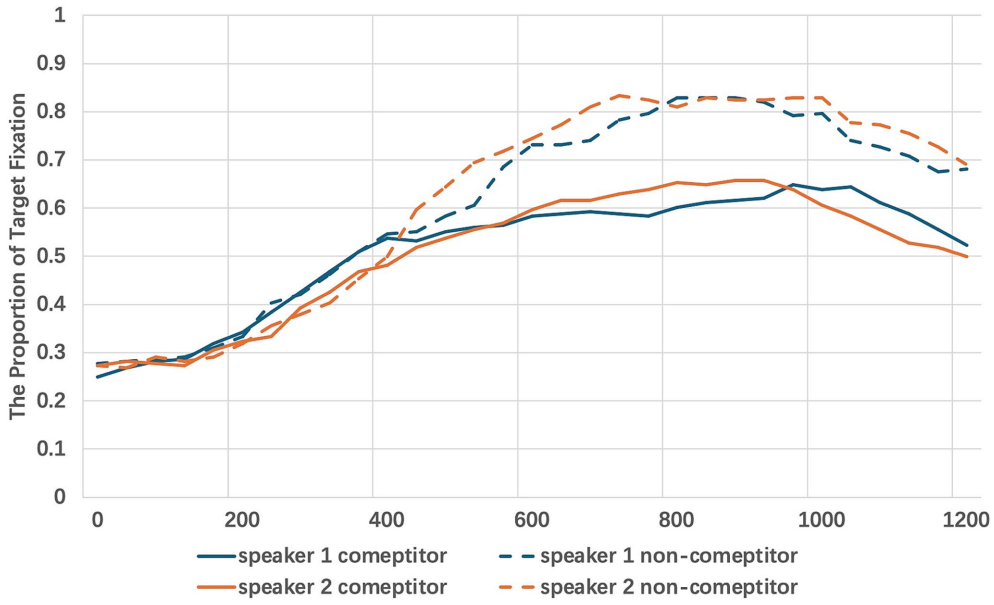


FIGURE 3
The 1200ms time window capturing participants' eye movements from the onset of the target object description across speakers and competitor types.

The results indicated a significant main effect of competitor type $\beta = 0.10$, SE = 0.03, t = 2.57, $p < 0.05$, suggesting that participants had significantly higher fixations to the target in the non-competitor condition compared to the competitor condition. The effect of speaker was not significant, $\beta = 0.01$, SE = 0.01, t = 0.19, $p = 0.84$, indicating no substantial difference in target fixations based on speaker identity. The interaction between speaker and competitor type was also not significant, $\beta = 0.01$, SE = 0.01, t = 0.71, $p = 0.47$, suggesting that the effect of competitor type on target fixation did not vary significantly across different speakers. Overall, the results exhibited patterns consistent with the behavioral data, highlighting a significant impact of competitor type on eye movements, while speaker identity and its interaction with competitor type did not significantly influence fixation patterns.

2.4 Discussion

In this experiment, participants collaborated with two confederates on an object-picking task. While all four candidate objects were visible to the participants, only three were visible to the confederate speakers. The results of this study indicate that, compared to the non-competitor condition, participants were influenced by their privileged knowledge and considered the unshared competitor objects more frequently. However, compared to the chance level, participants still tended to adopt the confederates' perspective and selected the shared target objects.

The results are consistent with predictions. Rationally, when the confederates referred to an object that had a duplicate competitor in the participants' privileged view, participants were expected to consider the confederates' perspective and exclude the competitor object from the referential domain. However, previous studies have also found that participants consider their privileged competitor object to some extent, given that the competitor is semantically consistent with the word label (Keysar et al., 2000). The effect of competitor types observed in this study supports this conclusion.

Given that few studies within this field of research have been entirely conducted in a Mandarin-speaking context, the present study provides exploratory results that could serve as a baseline for future studies. On the one hand, cross-cultural research suggests that in collectivist cultures like Chinese culture, people are more inclined to consider others' perspectives and focus more on the information in the common ground (Wu and Keysar, 2007b). On the other hand, Mandarin allows for bare nouns, which permit generic or category-level reference. Previous studies have shown that Mandarin bare nouns exhibit flexibility in interpretation, depending heavily on semantic and pragmatic cues in the discourse context (Kuo, 2008; Rullmann and You, 2006; Chang, 2016). That is, even if participants selected the unshared competitor objects, this selection could be considered acceptable. Unlike in English, where the determiner "the" in noun phrases provides additional cues for definite reference in common ground, bare nouns in Mandarin offer fewer such cues. Nevertheless, the results in this study are generally consistent with previous findings from English-language contexts, as participants' selection of the shared object was above chance level, demonstrating participants' perspective-taking process.

The next experiment will explore the main question of interest in this research: In perspective-taking, which is more significant: one's general stance or the specific shared information? In the following experiment, all experimental settings will remain the same, except that Confederate 1 will be able to see the objects in the blocked compartment through a gap between the partition. This means that Confederate 1 shares the same information with the participant, but they still share the same stance with Confederate 2 by seating together as speakers to guide the participant in completing the task. Should the results of the following experiment show the same trend as those of the first, exhibiting a similar gaze pattern during the competitor condition for both Confederate 1 and 2 trials, it would suggest that the general stance of the confederates may be influential. Conversely, if the results diverge from those of the first experiment, particularly with distinct processing of Confederate 1 and 2 trials, it would indicate a more significant role for shared knowledge inference. Furthermore, the shift in gaze pattern in Confederate 1 and 2 trials will also be valuable for further discussion, as it may offer deeper insights into the change in perspective-taking strategy when the similarity among the speakers and addressee changes.

3 Experiment 2

3.1 Participants

Twenty-one participants took part in Experiment 2 (11 female and 10 male, aged 17 to 23 years). The same two confederate speakers participated as in the first experiment. The recruitment and rewarding procedures were identical to those in Experiment 1. All participants are native Mandarin speakers, and none of the participants reported any cognitive, hearing, vision, or other related issues.

3.2 Methods

Experiment 2 used the same materials and procedure as Experiment 1. The only difference was that Confederate 1 could see the objects in the blocked compartment (see Figure 4). During the test, participants were first instructed to familiarise themselves with the shelf from the opposite side by sitting in the positions of C1 and C2 for several practice trials. In the formal test, the two confederate speakers took turns asking referential questions to the participants.

This experiment also consisted of 48 trials, including 16 non-competitor trials, 16 competitor trials, and 16 filler trials. Trials were evenly divided between C1 and C2, with each having 24 trials. The order of instructions was identical to that in Experiment 1, ensuring the two sets of experiments were comparable.

3.3 Results

3.3.1 Behavioral data analyses

In the second experiment, similar analyses were conducted to examine the effects of competitor types on target selection. As with the first experiment, Wilcoxon signed ranks tests were conducted separately for the trials in conditions C1 and C2, with competitor types (competitor vs. non-competitor) as the independent variable and target selection proportion as the dependent variable (Figure 2).

The results mirrored those of the first experiment. In C1 trials, participants selected the target item more frequently in non-competitor trials ($M = 1$, $SD = 0$) than in competitor trials ($M = 0.63$, $SD = 0.21$), $Z = -4.03$, $p < 0.001$, effect size = 1.762. Similarly, in condition C2, participants also selected the target item more often in non-competitor trials ($M = 1$, $SD = 0$) than in competitor trials ($M = 0.95$, $SD = 0.09$), $Z = -2.53$, $p < 0.05$, effect size = 0.556. This confirms that the presence of a competitor item significantly reduced participants' likelihood of selecting the target item, consistent with the results from the first experiment.

However, unlike in the first experiment, a paired samples Wilcoxon signed ranks test for the competitor trials revealed a significant difference between C1 and C2. Participants' target selection proportions in C1 were significantly lower than in C2, $Z = -3.94$, $p < 0.001$, effect size = 1.981. This suggests that whether the particular information is shared had a significant impact on how participants resolve the referential ambiguity.

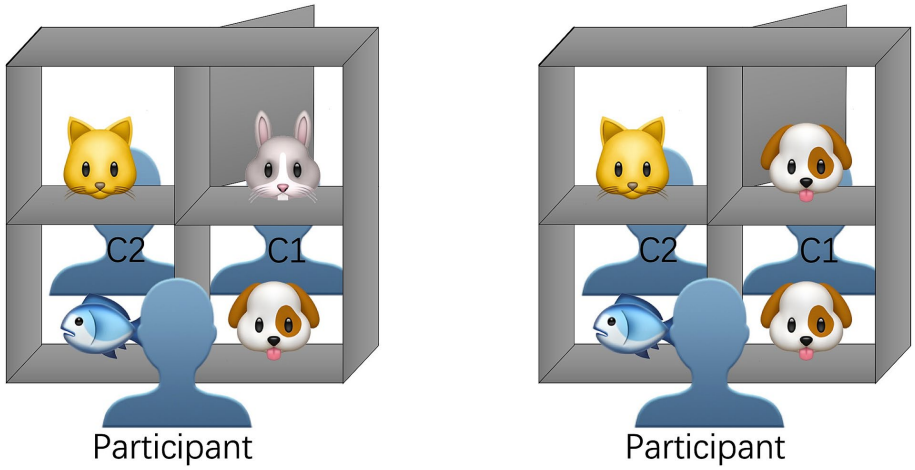


FIGURE 4
The display of the shelf and the seating arrangement of participants and confederates of Experiment 2.

Participants' target choice in the competitor condition was also compared to the chance level of 0.5. Two separate one-sample Wilcoxon signed ranks tests were performed for C1 and C2 trials. Results showed that for both C1 and C2, participants' target choice was significantly higher than chance, C1: $Z = 2.47$, $p < 0.05$, effect size = 0.619; C2: $Z = 4.17$, $p < 0.001$, effect size = 5, suggesting that participants' target choices in both C1 and C2 conditions in the competitor trials were significantly greater than the chance level of 0.5.

The results, on the one hand, suggest that participants refer to the shared information to constrain the referential domain, given the effect of speakers in the competitor trials. On the other hand, it also suggests that the general stance of the speakers influences how participants interpret the bare noun, given that even in C1 trials they chose the C2 shared target above chance level.

3.3.2 Combined analysis of Experiments 1 and 2

To further explore the effects of shared knowledge across both experiments, a combined analysis was conducted, focusing on the competitor conditions in C1 and C2 trials separately.

For C1, a Mann–Whitney U test revealed a significant difference between the competitor trials in Experiments 1 and 2, $Z = -2.440$, $p < 0.05$, effect size = 0.611. Participants selected the target item less frequently in Experiment 2, indicating a stronger effect of the competitor item when the confederate shared both referential candidates with the participants. For C2, no difference was found between Experiments 1 and 2 in competitor trials, $Z = -1.621$, $p = 0.105$, effect size = 0.727. This result is consistent with expectations, as C2's visual field remained unchanged, while in C1, both referential candidates were visible in Experiment 2. Since C2's position and visual field remained the same across both experiments, no difference across C2 was predicted. In the next section, participants' eye movements will be analysed to further explore their potential processing strategies.

3.3.3 Eye-tracking data analyses

As the first experiment, a general linear mixed-effects model was conducted using the lmer function in R to examine the impact of competitor type and speaker on participants' eye movements. The model included competitor types (competitor vs. non-competitor) and speaker (C1 vs. C2) as fixed effects, while participants and items were incorporated as random effects. The dependent variable was the fixation to target, measured in 40ms bins. The fixation towards the target was graphed over time in Figure 5.

The analysis revealed a significant main effect of competitor type, $\beta = 0.37$, SE = 0.05, t = 7.21, $p < 0.001$, indicating that fixations to the target were significantly higher in non-competitor trials compared to competitor trials. Additionally, a significant effect of speaker was observed, $\beta = 0.23$, SE = 0.03, t = 8.03, $p < 0.001$, suggesting that fixations to the target varied depending on the speaker's identity. Furthermore, a significant interaction between competitor type and speaker was found, $\beta = -0.19$, SE = 0.02, t = −12.764, $p < 0.001$, demonstrating that the effect of competitor type on target fixation differed across speakers.

To further explore the interaction between competitor type and speaker, post-hoc analyses were conducted using the emmeans package in R. The purpose was to clarify how the effect of competitor type on fixation to the target differed across speakers, as well as how the speaker effect varied between competitor and non-competitor conditions.

The analysis revealed significant differences between speakers in the competitor condition, estimate = −0.23, SE = 0.03, z = −8.04, $p < 0.001$, indicating that fixations to the target were significantly lower for C1 compared to C2 when competitors were present. In contrast, the comparison between speakers in the non-competitor condition was not significant, estimate = −0.04, SE = 0.03, z = −1.29, $p = 0.20$, suggesting that there were no notable differences in target fixations between C1 and C2 when no competitors were present.

The effect of competitor type was also analysed separately for each speaker. For C1, fixations to the target were significantly higher in the non-competitor condition compared to the competitor condition, estimate = −0.18, SE = 0.05, z = −3.81, $p < 0.001$. However, for C2, this effect was not significant, estimate = 0.02, SE = 0.05, z = 0.32, $p = 0.75$, indicating no substantial difference in target fixations between competitor and non-competitor conditions.

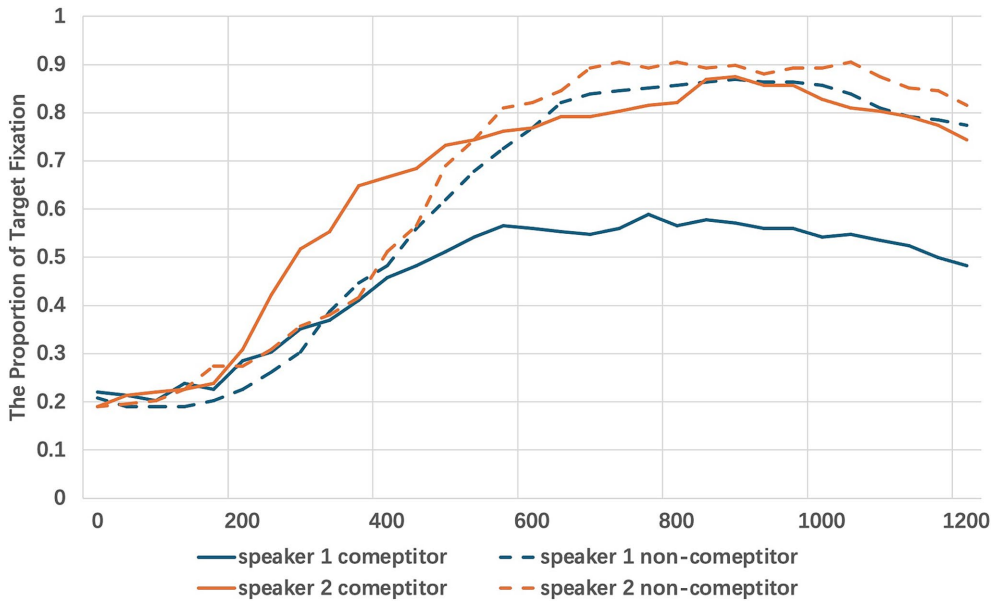


FIGURE 5
The 1200 ms time window capturing participants' eye movements from the onset of the target object description across different conditions and speakers.

3.3.4 Combined analysis of Experiments 1 and 2

A general linear mixed-effects model was conducted to investigate the effects of competitor type (competitor vs. non-competitor) and experiment (experiment 1 vs. experiment 2) on participants' eye movements, with separate analyses for each speaker. The model included competitor type and experiment as fixed effects, and participants and items as random effects. The dependent variable was fixation to the target, measured in 40 ms bins.

For C1, the model revealed significant effects of both experiment and competitor type. The effect of experiment was significant, $\beta = 0.05$, SE = 0.01, t = 5.28, $p < 0.001$, indicating that fixation patterns differed between experiment 1 and experiment 2. The effect of competitor type was also significant, $\beta = 0.24$, SE = 0.05, t = 4.37, $p < 0.001$, with greater fixations to the target in the non-competitor condition compared to the competitor condition. Moreover, the interaction between experiment and competitor type was significant, $\beta = -0.06$, SE = 0.01, t = −4.11, $p < 0.001$, suggesting that the effect of competitor type on target fixation varied across experiments.

Post-hoc contrasts indicated a significant difference in fixation to the target between experiment 1 and experiment 2 in the competitor condition, estimate = −0.059, SE = 0.01, z = −5.28, $p < 0.001$, with lower fixations in experiment 2. However, no significant difference was observed between experiments in the non-competitor condition, estimate = 0.01, SE = 0.01, z = 0.51, $p = 0.60$.

For C2, the analysis for speaker 2 also showed significant effects of experiment and competitor type. The effect of experiment was significant, $\beta = -0.17$, SE = 0.05, t = −3.49, $p = 0.001$, indicating that fixation patterns differed between experiments. Competitor type had a significant effect on fixation, $\beta = -0.15$, SE = 0.04, t = −3.50, $p = 0.004$, with higher fixations in the non-competitor condition. The interaction between experiment and competitor type was significant, $\beta = 0.13$, SE = 0.01, t = 9.76, $p < 0.001$, highlighting that the impact of competitor type on fixation varied between experiments.

Post-hoc contrasts revealed a significant difference in target fixation between experiment 1 and experiment 2 in the competitor condition, estimate = 0.17, SE = 0.05, z = 3.49, $p < 0.001$, with higher fixations in experiment 2. In contrast, no significant difference was found between experiments in the non-competitor condition, estimate = 0.04, SE = 0.05, z-ratio = 0.81, $p = 0.41$.

On the one hand, the eye-tracking analysis revealed the predicted result: participants were more influenced by the competitor objects when they were shared with Confederate 1 in Experiment 2. On the other hand, the unexpected finding was that changes in Confederate 1's visual field also affected how participants took the perspective of Confederate 2. In particular, the distinct visual field of Confederate 1 served as a reminder to participants that Confederate 2 was unaware of the objects in the blocked compartment. This implies that the comparison of knowledge between the two confederates highlights the importance of recognizing the common ground between the speakers and the participants.

3.4 Discussion

In Experiment 2, both confederates sat opposite the participant, acting as speakers and directing the participant in an object-picking task. However, the two confederates had different visual fields: one confederate (C1) shared all four pieces of object information with the participant, whereas the other confederate (C2) did not. Consequently, the two confederates had different levels of shared information. Therefore, this study revealed how participants adjusted their perspective-taking strategies based on their partners' roles and the information shared.

First, the results confirmed that participants interpreted each confederate's instructions separately, depending on the information they shared with that particular confederate. Behavioral data showed

that although participants tended to choose the shared target in both C1 and C2 trials, the target selection rate was significantly higher in C2 trials than in C1 trials. The eye-tracking data showed similar results: participants' fixations on the target were significantly higher in C2 trials than in C1 trials. Furthermore, the eye-tracking data revealed more implicit processing by participants. In C1 trials, participants exhibited a competitor effect: fixations on the target were significantly higher in the non-competitor condition compared to the competitor condition. However, this effect was absent in C2 trials. This was because, in C2 trials, a compartment obscured the other candidate object, and participants rarely considered this candidate, which was visible only to themselves. In summary, although confederate 1 and confederate 2 gave identical instructions under the same object arrangement context, participants did not process these instructions consistently. Instead, they engaged in *partner-specific* interpretation based on the shared information with each confederate.

Second, a comparison between two experiments revealed that participants' perspective-taking with confederate 1 also influenced their interpretation of confederate 2's instructions. Although behavioral data indicated that participants only exhibited differences in target selection for the competitor condition in C1 trials (and no differences in C2 trials), the eye-tracking data explained the implicit attentional differences underlying these choices. Specifically, in Experiment 2, participants had fewer fixations to the target in C1 trials compared to Experiment 1, but had more fixations to the target in C2 trials. This opposite pattern indicates that an increase in shared information with one confederate heightened participants' awareness of the lack of shared information with the other. This aligns with previous findings that show similarity with the interlocutor actually leads to less perspective-taking by participants. Since confederate 2 was a less similar interlocutor to the participants, compared to confederate 1, participants took confederate 2's perspective more effectively than in the baseline situation from Experiment 1, and it is possible that certain expectations are formed even before the conversation begins.

4 General discussion

This study investigated how native Mandarin speakers rely on perspective information for referential resolution. Research in this area within a Mandarin-speaking context is still relatively limited, so the first experiment in this study provides a baseline reference for future studies. By using Mandarin bare nouns, the critical noun could refer to the shared object, the blocked object, or both. This study offers insights into how participants may take perspectives to make referential resolution when non-target object referents under indefinite or generic interpretations are acceptable. The results from Experiment 1 showed that participants still used perspective information to interpret bare nouns in a definite manner at a level higher than chance.

Another goal of this study was to establish a conversational scenario to investigate whether participants would rely more on the shared information with each confederate speaker or on the confederates' role as speakers as a whole. The results supported the former: participants considered the shared information with each speaker individually. In Experiment 2, participants interpreted the instructions from confederate 1 and confederate 2 differently, demonstrating that they distinguished between the perspectives of the two speakers. Additionally, although the results revealed that participants did not view two confederates as a whole, they reinforced the contrast between the speakers' perspectives. In both experiments, although confederate 2 maintained the same position and visual field, participants' interpretation of her instructions varied. Specifically, competitor inference was reduced in Experiment 2, particularly in eye-tracking data participants exhibited shorter latencies to move their eyes to the target in the competitor condition, compared with the non-competitor conditions. The results conform to the proposal that people take a stronger perspective of the one who has greater difference from themselves.

Meanwhile, the present study offers some additional issues based on its manipulation. The following will illustrate two issues: visual perspective-taking and multi-party conversations.

4.1 Visual perspective-taking

In this study, participants are involved in visual perspective-taking or Level 1 Theory of Mind (ToM). Visual perspective-taking refers to the ability to understand what someone else can or cannot see from their viewpoint. Level 1 ToM is the ability to understand that others might not know what they know. The present study does not concern the difference between the two; for example, the neural mechanisms underlying both visual perspective-taking and theory of mind are overlapping yet distinct (Ogawa and Matsuyama, 2022). However, the present manipulation tested how people infer what others know based on how they can see. That is, the present study focuses on the overlapping parts between these two concepts. Visual perspective-taking or Level 1 ToM ability typically develops early in childhood, before the age of 4, at which point Level 2 Theory of Mind matures (Moll and Tomasello, 2006; Flavell et al., 1981). These abilities are not cognitively demanding, making them feasible for real-time conversational use. They are also closely tied to attention allocation. In a dot perspective-taking experiment, participants viewed images of a figure surrounded by walls with or without dots. When the participants' perspective aligned with the figure's, that is, the dots visible to the figure were also visible to the participants, judgments about the figure's perspective were significantly faster than when perspectives were misaligned, with participants able to see dots behind the figure but the figure unable to see them (Samson et al., 2010). Based on this, Heyes (2014) proposed that such processing might be more about attentional orienting than mentalising. When another's perspective conflicts with one's own, cognitive competition and inhibition occur between the two alternatives, similar to the Simon effect.

The results of this study contribute to research on visual perspective, specifically regarding how the consistency between orientation and visible information affects participants' referential resolution. In Experiment 1, both confederates sat opposite the participant and could not see the blocked objects. The orientation of their seating and the information visible in that direction were consistent. Results showed that participants interpreted both confederates' instructions similarly, demonstrating partial but significant perspective-taking. However, in Experiment 2, confederate

2 was still seated opposite the participant and could not see the blocked objects, and the alignment between her seating orientation and the visible information remained consistent. Yet, confederate 2's orientation only matched confederate 1's seating direction, rather than her visible information. Focusing on participants' understanding of Confederate 2's instructions, the results showed a change in participants' strategies. Specifically, in the second experiment, participants' understanding of Confederate 2's speech was less influenced by their privileged ground compared to the first experiment. This change might have been influenced by the difference in what Confederate 1 and Confederate 2 could see, or by the misalignment between Confederate 1's seating direction and the visible information. These findings suggest that judgments of visual perspective go beyond simply determining whether something is in or out of the visual field and may also involve an attentional adjustment process. During this process, the consistency of different partners, or the alignment between the visual field and the visual direction, may play a role.

4.2 Multiparty conversation

One notable difference between this study and previous visual perspective-taking research is the potential impact of the number of conversational partners. In one-to-one conversations with two interlocutors, participants can easily integrate both the seating direction and the visual field of the other. Participants can judge whether the direction their partner faces aligns with the visible information or simply assess whether the direction their partner faces aligns with their own. Such a process is less resource-intensive, and interlocutors can thus more easily exclude their privileged knowledge from the referential domain. However, as the number of collaborators increases, the situation becomes more complex. First, each partner cannot simultaneously occupy the same physical space, implying potential differences in visual perspectives. In this study, although confederates 1 and 2 sat on the same side of the shelf, their views could not be identical, and in Experiment 2, these differences were further amplified due to gaps between the partitions in the shelf. Second, when multiple conversational partners are involved, information can be shared by all, some, or none of the participants. With each additional partner, individuals not only need to consider the common ground between themselves and the new interlocutor, but they may also infer the common ground shared between the other two interlocutors. This process imposes a heavy cognitive load. Additionally, even when each individual's visual perspective is clear, it is not always evident whose perspective the other partners are taking. In summary, the complexity of perspective-taking in multiparty situations increases exponentially. Therefore, individuals may rely on diverse strategies, in addition to shared knowledge, to navigate multiparty conversations.

The results of this study offer several insights into multiparty conversation, particularly regarding how individuals integrate the perspectives of multiple conversational partners. One strategy is to separately analyse each partner's perspective. Although this method is cognitively demanding, it remains feasible when the number of participants is small. The findings from Experiment 2 support this strategy, as participants demonstrated differentiated understanding of confederate 1 and 2's instructions, showing that participants have sufficient cognitive resources to distinguish between the perspectives of different interlocutors.

Another strategy for managing multiple partners in conversation involves taking a unified perspective as the number of participants increases. This could be based on the perspective of one particular individual, such as the less knowledgeable person, or it could involve combining the perspectives of all participants at a certain level. Previous research has shown that when the group size is small, such as when only three individuals are present, speakers tend to communicate according to the perspective of the least knowledgeable person, a strategy known as "Aim Low" (Yoon and Brown-Schmidt, 2014, 2018). As the number of conversational partners increases, individuals may shift to a combined strategy, tailoring their expressions based on the group's collective knowledge (Yoon and Brown-Schmidt, 2019). Both strategies indicate that addressees either compare or combine the knowledge status of all their conversational partners. The current study's results support this approach to some extent, as participants compared the perspectives of multiple partners. Specifically, the reduced influence of competitor inference in confederate 2 trials in Experiment 2 suggests that participants engaged in comparison across all partners, leading to a more refined understanding of confederate 2's perspective.

The finding that the participants compare the perspectives of multiple conversational partners also parallels some previous findings. Studies have shown that speakers in group conversations strive to accommodate multiple perspectives, providing clarifications that might not be necessary in one-on-one conversations. Although group conversation is structurally more complex than one-to-one interactions, research has demonstrated that after interacting with multiple individuals, people's discourse becomes more comprehensible to others (Fay et al., 2000; Lev-Ari and Sebanz, 2020).

Lastly, but not least, this study offers evidence from the perspective of the addressee in multiparty conversation. Previous studies primarily focus on audience design from the speaker's perspective, examining how speakers tailor their utterances based on the presence of various listeners. However, it remains unclear whether addressees, as recipients of information, adjust their language comprehension similarly. On the one hand, comprehension and production can follow similar patterns. Participants may not only consider the speaker's perspective but also take into account the speaker's strategy, which might involve considering the perspectives of others. For example, if "Rockefeller Centre" is common ground for New Yorkers A and B but unfamiliar to C, B might interpret A's reference to a "building with flags of various nations" differently depending on whether C is present. If C is absent, B might think A is referring to another building. If C is present, B might infer that A is indeed referring to Rockefeller Centre. In this scenario, B's comprehension adapts not only based on C's perspective but also on A's communicative strategy. On the other hand, strategies for processing common ground during production and comprehension may differ. While the speaker directs discourse to multiple individuals, the addressee only processes information from one person at each turn. In this sense, perspective-taking for addressees in multiparty settings may resemble that in one-on-one conversations. The results of this study seem to support the latter, as

participants adjusted their perspective-taking strategies based on confederate 1 and 2's knowledge separately. However, it may be due to the small group size. Future studies can explore how addressees adjust their strategy to accommodate a larger amount of perspectives.

5 Conclusion

The present study, through two experiments, examined how participants utilised the perspective information of two speakers to perform referential resolution of nouns. First, this research provides evidence from a Mandarin context, particularly addressing the role of bare nouns, which do not require determiners in Mandarin. The results show that participants largely considered perspective information during referential resolution. Second, the study revealed that participants, as addressees, not only relied on distinct perspectives but also made comparisons between the speakers' perspectives. This finding aligns with previous production-based studies and offers further evidence for perspective-taking in multiparty conversations from a comprehension standpoint. Third, this research focused primarily on visual perspective and small conversational groups. Future studies could expand on this by exploring more complex forms of perspective information and larger conversational groups, which may involve even more dynamic processes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found a: https://osf.io/y3xcg/?view_only=fd8941f5f81f49299a5ec9cbc9a18def.

Ethics statement

The studies involving humans were approved by Ethics Committee in Shenzhen University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

XZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. CS: Conceptualization, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing, Data curation, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Peking University Humanities and Social Sciences Starting Fund.

Acknowledgments

We would like to express our heartfelt gratitude to our student assistants, Xiu Yu, Xiaopei Lin, Zhi Xia, and Xinyu Pan, for their invaluable contributions in recruiting participants and collecting data. Their dedication and hard work have been essential to the success of this research.

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

Aumann, R. J. (1976). Agreeing to disagree. *Ann. Stat.* 4, 236–1239. doi: 10.1214/aos/1176343654

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *J. Mem. Lang.* 61, 171–190. doi: 10.1016/j.jml.2009.04.003

Brown-Schmidt, S. (2012). Beyond common and privileged: gradient representations of common ground in real-time language use. *Lang. Cogn. Proc.* 27, 62–89. doi: 10.1080/01690965.2010.543363

Brown-Schmidt, S., Gunlogson, C., and Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition* 107, 1122–1134. doi: 10.1016/j.cognition.2007.11.005

Brown-Schmidt, S., and Hanna, J. E. (2011). Talking in another person's shoes: incremental perspective-taking in language processing. *Dial. Disc.* 2, 11–33. doi: 10.5087/dad.2011.102

Brown-Schmidt, S., Yoon, S. O., and Ryskin, R. A. (2015). People as contexts in conversation. *Psychol. Learn. Motiv.* 62, 59–99. doi: 10.1016/bs.plm.2014.09.003

Chambers, J. R., Epley, N., Savitsky, K., and Windschitl, P. D. (2008). Knowing too much: using private knowledge to predict how one is viewed by others. *Psychol. Sci.* 19, 542–548. doi: 10.1111/j.1467-9280.2008.02121.x

Chang, H. H. (2016). Interpretation of bare and demonstrative noun phrases in the acquisition of mandarin. *Lang. Linguist.* 17, 797–825. doi: 10.1177/1606822X16660952

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Clark, H. H., and Marshall, C. (1978). "Reference diaries" in Theoretical issues in natural language processing. ed. D. L. Waltz, vol. *2* (New York, New York: Association for Computing Machinery), 57–63.

Clark, H. H., and Marshall, C. R. (1981). "Definite reference and mutual knowledge" in Elements of discourse understanding. eds. A. K. Joshi, B. L. Webber and I. A. Sag (Cambridge: Cambridge University Press), 10–63.

Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.* 87, 327–339. doi: 10.1037/0022-3514.87.3.327

Fay, N., Garrod, S., and Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: the influence of group size. *Psychol. Sci.* 11, 481–486. doi: 10.1111/1467-9280.00292

Flavell, J. H., Everett, B. A., Croft, K., and Flavell, E. R. (1981). Young children's knowledge about visual perception: further evidence for the level 1–level 2 distinction. *Dev. Psychol.* 17, 99–103. doi: 10.1037/0012-1649.17.1.99

Hallett, P. E. (1986). "Eye movements" in Handbook of perception and human performance I: Sensory processes and perception. eds. K. B. Boff, L. Kaufman and J. P. Thomas (New York: Wiley), 10–102.

Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *J. Mem. Lang.* 49, 43–61. doi: 10.1016/S0749-596X(03)00022-6

Heller, D., and Brown-Schmidt, S. (2023). The multiple perspectives theory of mental states in communication. *Cogn. Sci.* 47:e13322. doi: 10.1111/cogs.13322

Heller, D., Gorman, K. S., and Tanenhaus, M. K. (2012). To name or to describe: shared knowledge affects referential form. *Top. Cogn. Sci.* 4, 290–305. doi: 10.1111/j.1756-8765.2012.01182.x

Heller, D., Parisien, C., and Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition* 149, 104–120. doi: 10.1016/j.cognition.2015.12.008

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspect. Psychol. Sci.* 9, 131–143. doi: 10.1177/1745691613518076

Horton, W. S., and Gerrig, R. J. (2005a). The impact of memory demands on audience design during language production. *Cognition* 96, 127–142. doi: 10.1016/j.cognition.2004.07.001

Horton, W. S., and Gerrig, R. J. (2005b). Conversational common ground and memory processes in language production. *Discourse Process.* 40, 1–35. doi: 10.1207/s15326950dp4001_1

Horton, W. S., and Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Top. Cogn. Sci.* 8, 780–795. doi: 10.1111/tops.12216

Huang, Y. T., Zheng, X., Meng, X., and Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of mandarin passive sentences. *J. Mem. Lang.* 69, 589–606. doi: 10.1016/j.jml.2013.08.002

Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychol. Sci.* 11, 32–38. doi: 10.1111/1467-9280.00211

Keysar, B., Barr, D. J., and Horton, W. S. (1998). The egocentric basis of language use: insights from a processing approach. *Curr. Dir. Psychol. Sci.* 7, 46–49. doi: 10.1111/1467-8721.ep13175613

Kronmüller, E., and Barr, D. J. (2007). Perspective-free pragmatics: broken precedents and the recovery-from-preemption hypothesis. *J. Mem. Lang.* 56, 436–455. doi: 10.1016/j.jml.2006.05.002

Kuo, J. Y. C. (2008). A pragmatic approach to the interpretations of mandarin bare nouns. *J. Pragmat.* 40, 1082–1102. doi: 10.1016/j.pragma.2008.03.004

Lev-Ari, S., and Sebanz, N. (2020). Interacting with multiple partners improves communication skills. *Cogn. Sci.* 44:e12836. doi: 10.1111/cogs.12836

Li, C. N., and Thompson, S. (1975). "The semantic function of word order: a case study in mandarin" in Word order and word order change. ed. C. N. Li (Austin: University of Texas Press), 163–186.

Li, C. N., and Thompson, S. A. (1989). Mandarin Chinese: a functional reference grammar. Berkeley and Los Angeles: University of California Press.

Metzing, C., and Brennan, S. E. (2003). When conceptual pacts are broken: partner-specific effects on the comprehension of referring expressions. *J. Mem. Lang.* 49, 201–213. doi: 10.1016/S0749-596X(03)00028-7

Moll, H., Carpenter, M., and Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Dev. Sci.* 10, 826–835. doi: 10.1111/j.1467-7687.2007.00615.x

Moll, H., Carpenter, M., and Tomasello, M. (2011). Social engagement leads 2-year-olds to overestimate others' knowledge. *Infancy* 16, 248–265. doi: 10.1111/j.1532-7078.2010.00044.x

Moll, H., and Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Dev.* 82, 661–673. doi: 10.1111/j.1467-8624.2010.01571.x

Moll, H., and Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *Br. J. Dev. Psychol.* 24, 603–613. doi: 10.1348/026151005X55370

Mozuraitis, M., Stevenson, S., and Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cogn. Sci.* 42, 974–1008. doi: 10.1111/cogs.12582

Nadig, A. S., and Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychol. Sci.* 13, 329–336. doi: 10.1111/j.0956-7976.2002.00460.x

Ogawa, K., and Matsuyama, Y. (2022). Heterogeneity of social cognition in temporo-parietal junction: overlapping yet distinct representation between visual perspective-taking and theory of mind. *bioRxiv*, 2022-01. doi: 10.1101/2022.01.04.474884

Pickering, M. J., and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* 27, 169–190. doi: 10.1017/S0140525X04000056

Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495

Pickering, M. J., and Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge: Cambridge University Press.

Robbins, J. M., and Krueger, J. I. (2005). Social projection to ingroups and outgroups: a review and meta-analysis. *Personal. Soc. Psychol. Rev.* 9, 32–47. doi: 10.1207/s15327957pspr0901_3

Rubin, R. D., Brown-Schmidt, S., Duff, M. C., Tranel, D., and Cohen, N. J. (2011). How do I remember that I know you know that I know? *Psychol. Sci.* 22, 1574–1582. doi: 10.1177/0956797611418245

Rullmann, H., and You, A. (2006). "General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese," in *Where semantics meets pragmatics*. eds. K. von Heusinger and K. P. Turner (Amsterdam: Elsevier), 175–196.

Runkel, P. J. (1956). Cognitive similarity in facilitating communication. *Sociometry* 19, 178–191. doi: 10.2307/2785631

Ryskin, R., Stevenson, S., and Heller, D. (2020). "Probabilistic weighting of perspectives in dyadic communication," in *Proceedings of the 42nd annual conference of the cognitive science society*. Toronto, CA. pp. 252–258.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., and Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1255–1266. doi: 10.1037/a0018729

Savitsky, K., Keysar, B., Epley, N., Carter, T., and Swanson, A. (2011). The closeness-communication bias: increased egocentrism among friends versus strangers. *J. Exp. Soc. Psychol.* 47, 269–273. doi: 10.1016/j.jesp.2010.09.005

Schiffer, S. R. (1972). Meaning. Oxford: Oxford University Press.

Sikos, L., Tomlinson, S. B., Heins, C., and Grodner, D. J. (2019). What do you know? ERP evidence for immediate use of common ground during online reference resolution. *Cognition* 182, 275–285. doi: 10.1016/j.cognition.2018.10.013

Stalnaker, R. C. (1978). "Assertion," in *Syntax and Semantics 9: Pragmatics*. ed. P. Cole (New York, NY: Academic Press), pp. 315–332.

Tomasello, M., and Farrar, M. J. (1986). Joint attention and early language. *Child Dev.* 57, 1454–1463. doi: 10.2307/1130423

Wu, S., and Keysar, B. (2007a). The effect of information overlap on communication effectiveness. *Cogn. Sci.* 31, 169–181. doi: 10.1080/03640210709336989

Wu, S., and Keysar, B. (2007b). The effect of culture on perspective taking. *Psychol. Sci.* 18, 600–606. doi: 10.1111/j.1467-9280.2007.01946.x

Yoon, S. O., and Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 919–937. doi: 10.1037/a0036161

Yoon, S. O., and Brown-Schmidt, S. (2018). Aim low: mechanisms of audience design in multiparty conversation. *Discourse Process.* 55, 566–592. doi: 10.1080/0163853X.2017.1286225

Yoon, S. O., and Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cogn. Sci.* 43:e12774. doi: 10.1111/cogs.12774

Zheng, X., and Breheny, R. (2021). Investigating the basis of memory-based effects on common ground. *Lang. Cogn. Neurosci.* 36, 941–958. doi: 10.1080/23273798.2021.1899254

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 06 February 2025
DOI 10.3389/fpsyg.2024.1494192

Check for updates

OPEN ACCESS

EDITED BY
Elena Nicoladis,
University of British Columbia, Canada

REVIEWED BY
Isabella Poggi,
Roma Tre University, Italy
Cornelia Müller,
European University Viadrina, Germany

*CORRESPONDENCE
Silva H. Ladewig
✉ silva.ladewig@uni-goettingen.de

RECEIVED 10 September 2024
ACCEPTED 10 December 2024
PUBLISHED 06 February 2025

CITATION
Ladewig SH (2025) Embodied sharpness: exploring the slicing gesture in political talk shows.
Front. Psychol. 15:1494192.
doi: 10.3389/fpsyg.2024.1494192

COPYRIGHT
© 2025 Ladewig. 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.

Embodied sharpness: exploring the slicing gesture in political talk shows

Silva H. Ladewig*

Department of German Philology, University of Göttingen, Göttingen, Germany

This paper explores the Slicing gesture within German political talk shows, focusing on its role in recurrent gesture sequences observed in German political talk shows. Through a detailed historical overview of recurrent gestures in political communication and an extension of the repertoire of recurrent gestures in German speakers, the study emphasizes the Slicing gesture's function in stance-taking and self-presentation. Highlighting its forms and functions, the analysis demonstrates how this gesture enacts sharpness, decisiveness, and clarity. The study integrates embodied and phenomenological perspectives, showing how kinesthetic experiences shape the gesture's meaning and metapragmatic dimensions. By contributing to the understanding of recurrent gestures as multimodal resources in political discourse, the paper sheds light on the interplay between embodied communication and rhetorical style.

KEYWORDS

recurrent gestures, slicing gesture, gesture sequences, metapragmatic meaning, kinesthetic experiences of gestures, stance taking in political communication, mediated multimodal communication, gesture stabilization

1 Introduction

Understanding how speakers encode and express points of view is a central topic in linguistics. According to Stubbs (1996, p. 202), "[a]ll utterances encode such a point of view, and the description of the markers of such points of view and their meanings should therefore be a central topic for linguistics." Stance-taking is an inherent part of interaction, as we inevitably adopt stances whenever we engage in communication (see Du Bois and Kärkkäinen, 2012). Du Bois (2007, p. 163) defines "stance as a public act by a social actor, achieved dialogically through overt communicative means (language, gesture, and other symbolic forms), through which social actors simultaneously evaluate objects, position subjects (themselves and others), and align with other subjects, with respect to any salient dimension of value in the sociocultural field." This perspective highlights that stance-taking is not merely a lexical phenomenon but a multifaceted activity involving active engagement by individuals. Consequently, stance is diverse and varied, yet consistently intertwined with the pragmatic and social aspects of human behavior. This understanding underscores the contextual, practical, and interactive nature of stance-taking (see Englebretson, 2007, pp. 2–3).

Bohmann and Ahlers (2022, p. 66) point out that while Du Bois' definition of stance as a "public act" is broad but generally accepted, its scope leaves the operationalization of stance in specific research contexts somewhat vague. They argue that this ambiguity affects both the formal aspect, which pertains to the nature of the "public act," and the functional aspect, which involves the variety and interrelationships of different stances taken during interaction. This paper will explore these broad concepts by focusing on how gestures, particularly the recurrent Slicing gesture, characterized by a flat hand with the edge moving downwards, fulfill various functions of stance-taking. By examining the kinesic and functional aspects of gestures in

interaction, this study aims to provide a clearer understanding of how different stances are expressed multimodally and related (juxtaposed or contrasted) through gesture sequences. Thus, this paper contributes to the growing body of research on stance-taking, expressed through various embodied means of communication such as gestures, head movements, and gaze (see Andries et al., 2023, for an overview) supporting the position that the concept of stance extends beyond verbal communication. In doing so, the study presented merges two aspects of multimodal, i.e., verbo-gestural, communication: recurrent gesture-based expression of stance and markers of social identity within the context of mediated political debates. Considering the first aspect, gesture studies have demonstrated that recurrent gestures, which are partly conventionalized gestures (examples are given in Figure 1), embody a speaker's attitude toward the discussed object of discourse (e.g., Teßendorf, 2014; Bressem and Müller, 2017). The Throwing away gesture (Bressem and Müller, 2017), for instance, can be used to downplay an argument made by the speaker, the Brushing aside gesture (Teßendorf, 2014) may dismiss a point, and the Stretched index finger (Bressem and Müller, 2014) may highlight a crucial aspect.

Regarding the second aspect, works on social action and role inhabitance (e.g., Silverstein, 2004; Agha, 2006) have shown, that stances are often "indexical of enregistered identities" (Kiesling, 2022, p. 412). Thus, language, particularly speech styles, indexes stances that are simultaneously recognized as having the potential to shape identities (Johnstone, 2017; Kiesling, 2022). These styles are collections of linguistic forms that are associated with social norms, values and beliefs linked to specific ideologies that are familiar to members of a speech community. These linguistic forms circulate within that particular community. It will be shown that the gesture under investigation belongs to the repertoire of speakers engaged in public political communication. This leads to the final aspect explored in this paper, specifically the context of mediated political debates. It is demonstrated that the Slicing gesture is particularly employed by participants in televised political debates to enact stance and embody the rhetorical qualities of sharpness, decisiveness and commitment. To investigate this aspect, the paper focuses on a distinctive feature of political talk shows: the use of extended sequences of recurrent gestures. As will be shown, the Slicing gesture is frequently embedded within gesture sequences that include different variants of this gesture and/or other recurrent gestures. The focus on gesture sequences highlights another way of expressing stance. Stance can be conveyed through linguistic elements, such as the epistemic phrase *I think* as well as through the sequential occurrence of stance markers (Kärkkäinen, 2003, 2006). Lempert (2008, p. 572) referencing Jakobson's (1960) poetic function, notes that "linearly co-occurring elements can be analyzed for their comparability, contrast and complementarity, likeness and unlikeness." These occurring structures can "diagrammatically motivate pragmatic effects – including stance effects" by organizing the signs they consist of into comparable, sequentially ordered units (ibid.). The sequentially and comparability of co-occurring elements in gesture sequences offer a rich area for analyzing how gestures complement and contrast each other to convey nuanced stances together with speech.

The paper is structured as follows: It begins with an exploration of recurrent gestures in political communication, highlighting their role in presenting arguments and reinforcing rhetorical and social styles. The paper then delves into the specific dimension of taking stance with the recurrent Slicing gesture, discussing how it operates on single verbal elements within gesture sequences and develops meta-pragmatic meaning. Subsequently, the focus shifts to the kinesthetic experiences fundamental to the meaning of the Slicing gesture. The

FIGURE 1
Repertoire of recurrent gestures determined in televised political debates data. The repertoire is based on Bressem and Müller's (2014) repertoire of recurrent gestures determined for adult German speakers. New candidates, as well as the counts for all identified recurrent gestures in this study, are highlighted in blue.

paper concludes with a discussion that synthesizes the findings and implications of the study.

2 Recurrent gestures in political communication

Recurrent gestures are characterized by their repeated occurrence with the same form and meaning across various communication contexts and among different speakers (Ladewig, 2014b, p. 31; Müller, 2017; Harrison and Ladewig, 2021; Ladewig, 2024). They are often derived from practical actions and are engaged in semantic and pragmatic meaning-making. In fact, the communicative potential of these kinds of gestures has always been a subject of study in the field of rhetoric (Ott, 1902; Mosher, 1916; Quintilian, 1969). Recurrent gestures, like any other gestures, are performed spontaneously, meaning their performance is not planned by a speaker. However, they differ from "singular gestures" (Müller, 2010, 2017), whose forms and meanings emerge while speaking. In the field of gesture studies, singular gestures are also called iconic gestures or metaphoric gestures (McNeill, 1992), "depictive gestures" (Streeck, 2009) or "representational gestures" (Kita, 2000). They have not undergone stabilization processes and are therefore not culturally shared as is the case with recurrent gestures and emblems.

Apart from getting insights into stabilization processes in communicative resources other than spoken or signed languages, recurrent gestures are an interesting research phenomenon because they are engaged in pragmatic meaning making. This aspect was first addressed from the point of view of rhetoric (Quintilian, 1969) and in the education of actors (Ott, 1902; Mosher, 1916). Following Kendon (2004, pp. 158–159) and other researchers, recurrent gestures can fulfil the following functions: (a) *modal*, i.e., framing how an utterance should be interpreted, (b) *performative*, i.e., enacting a speech act such as offering ideas or stopping someone, (c) *parsing*, i.e., punctuating the spoken discourse into logical components, and (d) *interactive and interpersonal functions* regulating turns at talk such as holding the floor or requesting a turn. These pragmatic functions make recurrent gestures a semiotic resource par excellence in the field of political communication, where the focus is on discussing arguments and sound reasoning. This hypothesis is corroborated by the study presented, which shows that 81% of gesture use consists of recurrent gestures in a five-hour data corpus of political talk shows (Table 1). In fact, recurrent gestures have already been described as a public event

TABLE 1 Distribution of gesture types in the data.

Recurrent gestures	2,747
Hybrid recurrent gestures	89
New candidates for recurrent gestures	158
Recurrent gestures in total	2,994
Pointing gestures	237
Singular gesture	281
Unclear cases	104
Emblematic gestures	34
Beat gestures	8
Gestures in total	3,658

over more than 2000 years ago. As Müller points out '[g]estures have been considered a public phenomenon over 2000 years of European gesture study, and have been analyzed, stylized, and taught by scholars primarily for their impact on an audience' (Müller, 1998, p. 30, translation S.L.). The description and training of gestures of speakers goes back to Cicero and Quintilian who regarded speech-accompanying gestures as worthy of cultivation (Müller, 1998; Kendon, 2004). Quintilian (1969) distinguishes between gestures which "naturally proceed from us simultaneously with our words" and "others which indicate things by means of mimicry" (XI, III.89). Almost all of the gestures Quintilian lists either mark speech acts, such as "demand, promise, summon, dismiss, threaten, supplicate [...] question or deny" (XI, III.88), relate to parts of speech, or express emotions and attitudes, such as "indicating joy, sorrow, hesitation, confession, or penitence" (XI, III.88). The forms of the gestures he meticulously describes are reminiscent of the recurrent gestures identified in modern gesture studies, which have pragmatic functions and are integral to verbal utterances. Among the gestural forms described are the Palm up gesture (Kendon, 2004; Müller, 2004), the Ring gesture (Kendon, 2004; Neumann, 2004), the Fist (Bressem and Müller, 2014) or the Finger bunch (Kendon, 1995, 2004).

More recent studies of gestures in the field of political communication explore the role of speakers' gestures and the attitudes they express in self-presentation, placing greater emphasis on this aspect over the purely interactive functions of gestures. Streeck's (2008) study of the gestures used by Democrats in the 2004 U.S. election campaign ties in with Quintilian's work. He documents a repertoire of recurrent gestures used by the members of the Democratic Party which, according to him, appeared to be reminiscent of a shared code or a "public gesture style" (Streeck, 2008, pp. 156, 178). These gestures primarily serve the function of visualizing the information structure of speech and thus help processing information. The gestures identified show "a surprising congruence between the type of gestures that Quintilian advocated" and "what appears to be an unspoken consensus about adequate gesticulation among the Democratic Party politicians" (Streeck, 2008, p. 178). One motivation for the emergence of such a way of gesturing is that the politicians are eager to keep a rhetorical style which may go along with the creation of persona in para-(social) interaction. Based on these observations Streeck argues, that the study of recurrent gestures in the field of political discourse may provide insights into the "theory of self-presentation (…) within the context of electoral politics" (Streeck, 2008, p. 183 referring to Goffman, 1956). This aspect is particularly relevant to the study presented in Section 3.2.2, where it is argued that the Slicing gesture develops the metapragmatic meaning of defining things with clarity and sharpness within extended gesture sequences. Consequently, the Slicing gesture can be considered a significant semiotic resource for presenting oneself with distinct rhetorical qualities.

Another compelling analysis of the use of recurrent gestures was conducted by Lempert (2011). He investigated Barack Obama's use of the Ring gesture (Kendon, 2004; Neumann, 2004; Müller, 2014a) in his Senate race (2004), the primary debates (2007–2008) and the presidential debates against the Republican candidate, John McCain. Interestingly, Lempert noted a shift in the first presidential debate where Obama appeared to stage himself as a sharp speaker not only verbally but also gesturally. The Ring gesture (referred to by Lempert as precision grip) became a semiotic resource for Obama to not only

make a sharp point but to present himself as "*being* sharp" (Lempert, 2011, p. 245, italics in the original). Thus, in a process of "reflexive reanalysis and conventionalization" (Lempert, 2011, p. 258), the Ring gesture has acquired a higher-order indexical value of being a sharp speaker which presupposes the lower-order focus of foregrounding a discursive object in his speech. In other words, the Ring has moved from focusing something *for* an addressee to making a sharp point *against* an addressee (ibid). "In semiotic parlance it may be termed a metapragmatic icon, to the extent that it reflexively (hence 'meta-') typifies communicative behavior as a social act ('pragmatic'), and does so by means of felt resemblance ('icon'-icity)" (Lempert, 2011, p. 258). Lempert notes that this process goes along with an indexical shift from working locally on parts of speech to pointing to a (candidate) brand that evolves interdiscursively. "The relevant units of analysis change as one moves through these orders as well, for the conditions under which 'brand' becomes recognizable, for instance, are quite different from those that motivate readings of speaker-persona" (Lempert, 2011, p. 262). In other words, the latter may become recognizable in the recurring uses of a gesture within one discourse where it not only visualizes recurring themes but also exhibits rhetoric qualities of the speaker. The former becomes possible in the relational field of political discourse (or of competition, as Lempert states, p. 259). Hence the "indexical icon of brand qualia" is an "interdiscursive precipitate" (Lempert, 2011, p. 245) that evolves over many political appearances.

To conclude, recurrent gestures can frame utterances, enact speech acts, punctuate discourse, and regulate interaction, making them indispensable in political debates for presenting arguments and reinforcing rhetorical and social styles of speaking and gesturing. The Slicing gesture, in particular, embodies decisiveness and clarity, highlighting its role as a semiotic resource in defining and emphasizing discourse positions. It will be explored in the following section.

3 Dimensions of taking stance with the slicing gesture

The Slicing gesture is the most common gesture in the data underlying this study. To be more precise in 5 h of mediated political discourse, i.e., German televised political talk shows, the gesture occurred 653 times over 27 speakers. However, this does not mean that all the speakers investigated are politicians. On the contrary, while four of the speakers are politicians, the rest are journalists and experts in political communication. What unites them in the data of televised talk shows is their engagement in publicly positioning themselves.

In the data, 3,658 gestures were documented, of which 2,994 are classified as recurrent gestures out of which 158 gestures are considered as candidates for the repertoire of recurrent gestures shown in Figure 1 but have not yet been included. Within these 5 h, we also recorded over 281 singular or depictive gestures, 237 pointing gestures, 34 emblematic gestures, and eight beat gestures. Additionally, 104 gestures were difficult to categorize and thus labeled as unclear (Table 1).

Figure 1 lists the distribution of different recurrent gestures over the data. For recurrent gestures already documented for German adults (Müller et al., 2013; Bressem and Müller, 2014), the most common gesture observed was the Palm up open hand (PUOH), occurring 619 times, followed by the Ring gesture (Kendon, 1995, 2004; Neumann, 2004) with 222 instances. The Cyclic gesture (Ladewig, 2014a) was noted 147 times, and the Holding away gesture appeared 122 times. The Stretched index finger was recorded 91 times, while the Fist gesture was identified 67 times. Sweeping away gestures (see also Kendon, 2004; Harrison, 2018, for English) occurred 64 times, Back and Forth gestures were observed 27 times, and Change gestures appeared 15 times. Brushing away gestures were recorded 14 times, Weighing up gestures were noted 18 times, Vague gestures were observed 8 times, Dropping of the hand gestures occurred seven times, and the Stretched index finger, moved horizontally, was recorded once.

For new recurrent gestures added to the repertoire, the Slicing gesture was the most common, observed 653 times. The Directing gesture (see Kendon, 2004, for English; Fricke, 2010, for German) appeared 292 times, while the Placing gesture (see also Tutton, 2011, for French and English) was noted 129 times. The Pressed thumb gesture was recorded 98 times, the Finger Bunch (see Kendon, 2004, for Italian speech) appeared 51 times, the Grip gesture (see Streeck, 2017, for English) was identified 44 times, and the Holding down gesture occurred 40 times.

Among the 89 hybrid recurrent gestures (Table 1), the following distribution was observed (Table 2): Directing combined with Palm up open hand occurred 46 times, the Holding away gesture combined with Sweeping away was noted 11 times, and the Bunch combined with Cyclic gesture was observed eight times. Additionally, the Slicing gesture combined with the Palm up open hand was documented eight times, the Slicing gesture combined with Directing appeared seven times, and the Pressed thumb combined with Cyclic gesture occurred three times. Furthermore, the Ring gesture combined with the Sweeping away gesture was observed 2 times, the Bunch combined with the Pointing gesture was noted once, the Slicing gesture combined with the Cyclic gesture was identified once, and the Slicing gesture combined with the Pointing gesture was recorded once.

The gesture under investigation has been previously recognized in studies of adult French speakers (Calbris, 2003, 2011, calling it Cutting gesture, Rigid Hand or the both-handed Frame gesture) and English speakers (Streeck, 2008 calling it "the Slice"; Lempert, 2017; Harrison, 2018 calling it "chopping"). While Calbris, Streeck, and Lempert documented many instances of this gesture in the context of public and political communication, Harrison only documented it once in his data corpus of non-public conversations.[1] Morris (1994) also

TABLE 2 Distribution of hybrid recurrent gestures in the data.

Directing + Palm up open hand	46
Holding away + Sweeping away	11
Bunch + Cyclic gesture	8
Slicing gesture + Palm up open hand	8
Slicing gesture + Directing	7
Pressed thumb + Cyclic gesture	3
Ring gestures + Sweeping away	2
Bunch + Pointing gesture	1
Slicing gesture + Cyclic gesture	1
Slicing gesture + Pointing gesture	1

1 Personal communication.

documented one variant of this gesture, referring to it as the Hand Chop, which is used to cut through an argument. He describes the most conventionalized variant of this gesture where one "stiff hand chops down on the upturned palm of the other hand" (p. 103). This variant was documented nine times in the data presented. According to Morris, the derivational base of this gestural form lies in miming the action of cutting downwards with an axe or imitating a karate chop. He observes that this gesture is often employed in heated debates when someone wants to make a strong and clear point. Additionally, Morris notes a "pragmatic affinity" (Lempert, 2017) with the clenched fist slammed down, "but the chopping action reflects a mood of greater precision" (Morris, 1994, p. 103).

3.1 Exploring the slicing gesture: quantitative analysis of its forms and functions

The investigation of the Slicing gesture proceeded as follows. First, a data corpus was established and all gestures were identified and annotated using the annotation software ELAN (Wittenburg et al., 2006). The frequencies of all recurrent gestures, both previously documented (Müller et al., 2013; Bressem and Müller, 2014) and undocumented, were subsequently determined resulting in the repertoire shown in Figure 1. As the Slicing gesture emerged as the most frequently occurring gesture in the data corpus that had not been previously documented for German spoken language before, it was subjected to a more thorough analysis. This involved annotating the gesture with regard to its forms and functions in relation to speech, following the Linguistic Annotation System of gestures (LASG, Bressem et al., 2013). Based on that the core form and form variants of the gesture were determined. In the next step, the gesture's functions were specified in relation to stance-taking categories, as the gesture was frequently embedded in contexts of positioning. The results of the analysis are as follows.

The core form of the Slicing gesture is characterized by an open hand often with the palm facing toward the speaker's body (117 cases, Figure 2A), facing diagonally the speaker's central gesture space (114 cases, Figure 2B), or being aligned with the sagittal plane (338 cases, Figure 2C). Other less frequent variants include a two-handed Slicing gesture, with the palms facing the speaker's central gesture space, as shown in Figure 2C, but directed to the left or right (54 occurrences). Orientations that appeared fewer than five times were categorized under "other orientations," with a total of 30 occurrences.

Another form parameter contributing to the core form is the movement direction, which is downward in a straight or arced pattern. The gesture "evokes the act of cutting something into pieces or slices. The edge of the hand laterally cuts up sections of space "as Calbris (2003, p. 31) propose. The core meaning of Slicing and thus singling out an entity has emerged from the action of slicing in which the hand symbolizes a blade or is a manual blade. This aspect is illuminated in Section 4.

For the Slicing gesture used in televised political debates, the functions shown in Figure 3 were determined, with the most frequently observed being the definition of a discourse object, accounting for 339 occurrences (51.9% of the cases). Discursive objects can include ideas or arguments with which the speaker positions themselves in the discourse (see Section 3.2.1). The discursive function, involving the intensification of arguments through emphasis, was observed in 112 occurrences (17.2% of the cases). The metapragmatic function, where the gesture not only defines discursive objects but also embodies the speaker's rhetorical act of clearly defining things for co-participants, appeared in 74 occurrences (11.3% of the cases) (see Section 3.2.2). A combined function of defining a discourse object and adding emphasis was present in 62 occurrences (9.5% of the cases). The metacommunicative (pragmatic modal) function, used to mark information structure or engage in speech activities such as enumeration, accounted for 43 occurrences (6.6% of the cases).

The performative function, where the gesture is used to perform speech acts, was documented in only 9 occurrences (1.4% of the cases). Other functions, including hesitation or substitution of speech, also accounted for 9 occurrences (1.4% of the cases). Some functions could not be clearly determined and were marked as unclear, constituting 5 occurrences (0.7% of all cases).

Considering the orientation of the palm, identified as the feature defining form variants, the following distribution was observed (Figure 4). All orientations predominantly convey the meaning of defining a discourse object, with the highest percentages observed in the orientations palm lateral towards body (PLTB, Figure 2A), and palm lateral diagonal towards center (PLdiTC, Figure 2B). The second most common function across all orientations is the discursive function, followed by the

FIGURE 2
Variants of the Slicing gesture documented in the data of German political talk shows: **(a)** the Slicing gesture oriented towards the speaker's body (PLTB), **(b)** the Slicing gesture oriented diagonally towards the speaker's central gesture space (PLdiTC), and **(c)** the Slicing gesture aligned with the sagittal plane (PLTC).

Dominant functions	Number	Percentage
Define a discourse object	339	51.9%
Discursive function	112	17.2%
Metapragmatic function	74	11.3%
Define a discourse object + emphasis	62	9.5%
Metacommunicative function	43	6.6%
Performative function	9	1.4%
Other functions (incl. hesitation, speech-replaicing function)	9	1.4%
Unclear cases	5	0.7%
Total	653	100%

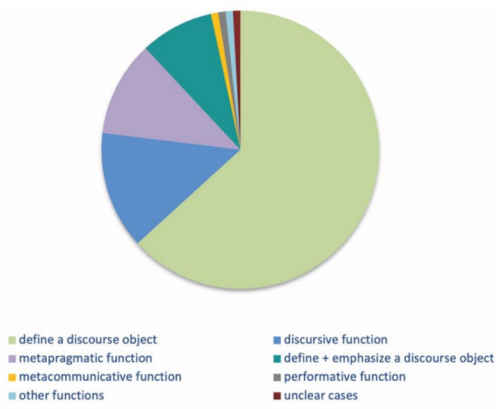


FIGURE 3
Most common functions of the Slicing gesture used in German televised political debates.

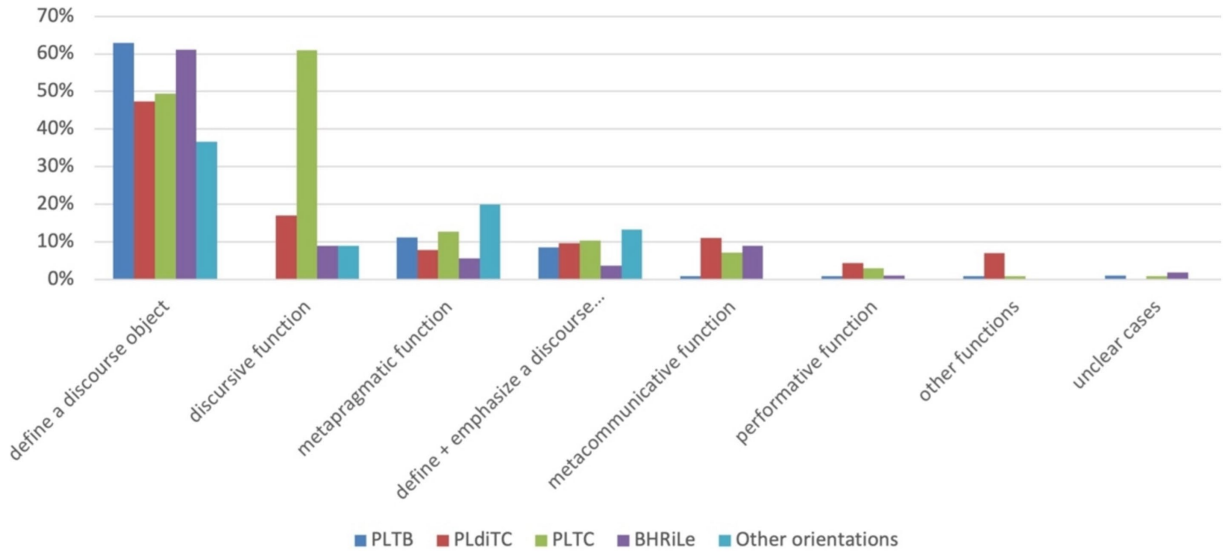


FIGURE 4
Distribution of functions by palm orientation. PLTB, palm lateral towards body (Figure 2A); PLdiTC, palm lateral diagonal towards center (Figure 2B); PLTC, palm lateral towards center (Figure 2C). The coding categories are based on Bressem (2013). BHRiLe, both hands facing each other, directing to the right or left with the palms in the sagittal plane.

metapragmatic function, particularly prevalent in the BHRiLe orientation (both hands facing each other, directing to the right or left with the palms in the sagittal plane). The combined function of defining a discourse object and emphasizing it is also notable in several orientations, while the metacommunicative and performative functions occur less frequently across all palm orientations.

One of the fundamental features of language is its capacity to concurrently depict subjects, objects, or events and convey a stance or perspective regarding these depictions (Andries et al., 2023). Similarly, gestures are multifunctional (e.g., Müller, 1998; McNeill, 2005; Kok, 2016), with the Slicing gesture exemplifying this versatility. This gesture indicates that speakers frequently engage in stance-taking beyond the functions already discussed. As Kiesling (2022, p. 417) notes, "stance is not simply a single utterance but rather emerges across multiple utterances." This highlights how both language and gestures work together to express complex meanings and stances in communication. Consequently, a speaker may define a discourse object while evaluating it or positioning themselves regarding the topic under discussion. The speaker may also frame (or define) a question that invites stance-taking from other participants. Alternatively, the gesture might be used to highlight a crucial distinction within a conversation, effectively separating two perspectives. By emphasizing one perspective and diminishing the other, the speaker can position themselves in the discussion, drawing attention to the preferred viewpoint while sidelining the less favored one. For instance, when distinguishing between two competing theories, the Slicing gesture can underscore the speaker's preference for one over the other by moving the hand

toward the speaker's body, subtly encouraging others to consider their stance. Thus, while addressing individual ideas verbally, the Slicing gesture simultaneously provide the marks these distinctions gesturally, providing a visual and physical representation of the separation and emphasis of different concepts.

Overall, the Slicing gesture is involved in stance-taking speech activity in 61% of the cases, equating to 401 out of 653 gesture occurrences. These cases can be further defined according to the different categories determined for stance taking (Figure 5).

Evaluation emerges as the most frequent function across all palm orientations, particularly in the palm lateral towards center orientation (PLTC, Figure 2C). This indicates that speakers commonly use the Slicing gesture to express evaluative stances, reflecting judgments or assessments about the discussed topic. The defining of one's own position is also a prominent function, especially visible in orientations like palm lateral towards body (PLTB, Figure 2A) and palm lateral directed towards center (PLTC, Figure 2C). These orientations are frequently employed by speakers to assert personal stances that align with a group to which the speakers belong, such as a political party or professional group.

Deontic stance, which involves conveying obligations, permissions, and prohibitions, is consistently observed across orientations, though it is less frequent than evaluation. Affective stance, while less common, is still present across multiple orientations, indicating the gesture's role in expressing emotions or attitudes.

Stance questions and intensification functions appear less frequently but are still noteworthy. Stance questions involve inquiring about another's viewpoint, while intensification involves emphasizing particular points. Both are present across orientations in varying degrees.

Finally, other stance-related functions, such as distancing, are observed but are the least frequent across all palm orientations.

Overall, the Slicing gesture serves multiple functions in stance taking, with evaluation being the most prevalent across various orientations, followed by defining one's own position and deontic stance. In what follows the focus of the paper will be shifted towards the qualitative analysis of the Slicing gesture, providing examples for some of the stance functions presented and highlighting different kinesic complexities and levels of gesture speech interaction.

3.2 Taking a stance with the slicing gesture: a qualitative analysis

Taking a stance through the Slicing gesture can manifest in various ways, depending on how it interacts with speech. In some cases, the gesture's meaning is closely tied to specific verbal elements, emphasizing particular words or phrases and reinforcing the speaker's point. In other instances, the Slicing gesture takes on a broader metapragmatic role, especially when used in extended gesture sequences. Here, it goes beyond individual utterances, embodying an overarching communicative strategy that helps define and clarify the speaker's stance throughout the discourse. This section explores these different dimensions of stance-taking through qualitative analysis, revealing the versatility and impact of the Slicing gesture.

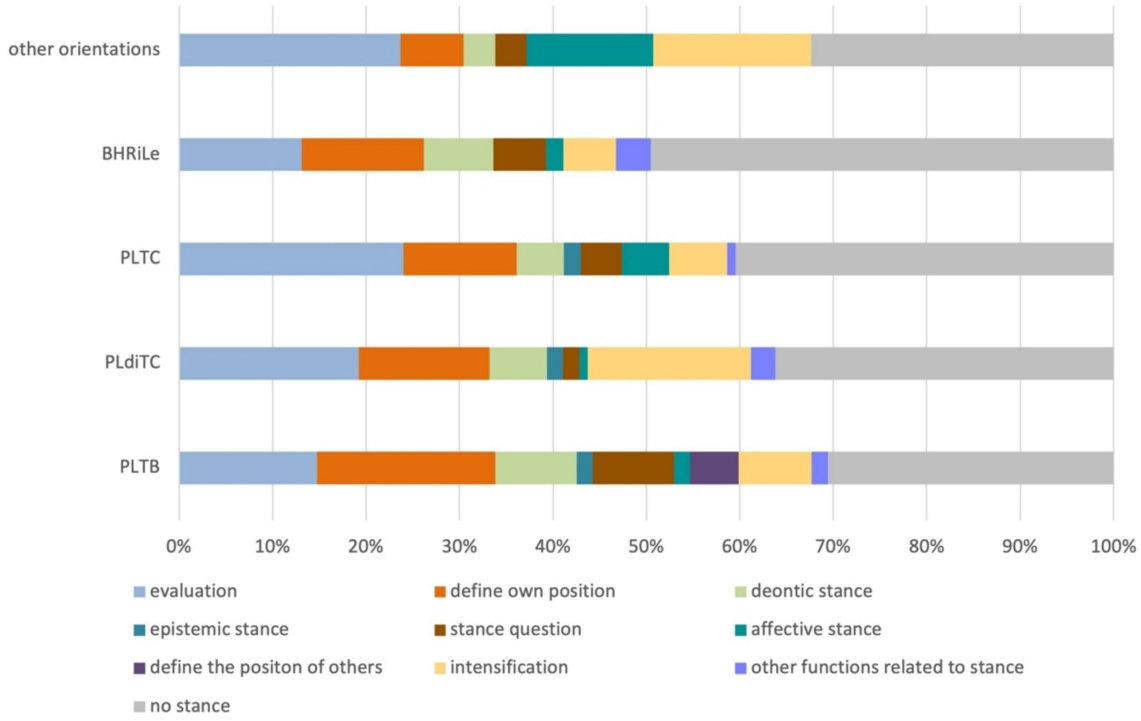


FIGURE 5
Functions of stance taking determined for the Slicing gesture and distribution over form variants. PLTB, palm lateral toward body (Figure 2A); PLdiTC, palm lateral diagonal towards center (Figure 2B); PLTC, palm lateral towards center (Figure 2C); BHRiLe, both hands facing each other, directing to the right or left.

Building on the diverse ways the Slicing gesture can convey meaning, what stands out is its frequent embedding in gesture sequences. Specifically, it often appears (a) in sequences of varying recurrent gestures or (b) in sequences involving different variants of the Slicing gesture. This pattern seems to be unique to the mediated contexts observed in this study, as such extended gesture sequences are rarely found in private settings.[2] Despite their prevalence in public discourse, gesture sequences remain an underexplored area in gesture studies. Exceptions include studies on recurrent gestures marking the topic comment structure of an utterance. Gesture sequences of this kind often show an "open-to" structure such as the a combination of a Ring gesture combined with an open Pistol hand (Seyfeddinipur, 2004) or the finger bunch or *grappolo* combined with a Palm up open hand (Kendon, 2004). In both cases, the sequences start with a specific handshape and movement, the Ring or the Finger bunch, marking the topic, and are followed by an opening of the hands as a comment on the topic given.

Harrison (2018) gives a more comprehensive account of sequences of recurrent gestures associated with negations. He defines these sequences as patterns of two gesture phrases within one gesture unit where the transition of the gestures involves the rotation of the wrist. The observed "sequences occur over utterances that ensue as part of a single rhetorical move, speech act, or argument, usually within the confines of a single turn at talk" (Harrison, 2018, p. 105). Moreover, Harrison observed that the rotation of the wrist reflects shifts at the level of discourse. Examples include the "Palm Up to Horizontal Palm Gesture Sequence" (Harrison, 2018, p. 108) which (a) embodies a topic comment structure, (b) is part of a conditional statement in which the gestures accompany the information structure of such statements, or (b) relates to a verbal context that precedes the gestures sequence while the gestures occur in absence of speech.

Gesture sequences are not only composed of different recurrent gestures but can also be formed by the repeated use of recurrent gestures. Some examples are given by Calbris (2003) who suggests that the repeated use of the Cutting gesture in French marks "the successive consideration of elements as they are cut out "(p. 43) or the singling out of objects (p. 33). Müller (2004) noted that the repeated downward movement of the Palm up open hand can be used to list a series of arguments. Similarly, Bressem (2021) documented the repetitive use of recurrent gestures that have a prosodic function.

The behavior of the Slicing gesture in gesture sequences will be explored in the following sections, with a focus on the process of multimodal meaning-making. Particular attention will be given to the various dimensions of stance-taking involving the Slicing gesture in sequences of its different variants, of different recurrent gestures, and different gesture types. Additionally, the Slicing gesture's conveyance of meta-pragmatic meaning in extended gesture sequences will be examined.

3.2.1 Operating on single verbal elements within gesture sequence of recurrent gestures

This section mainly focuses on examples in which the Slicing gesture operates on single verbal constructions of an utterance. This means that the gesture is closely linked with specific words or phrases within a sentence, thereby reinforcing the speaker's point. The chapter is organized based on the different gesture sequences observed in the data, including sequences of various variants of the Slicing gesture, sequences of different recurrent gestures, and sequences of different gesture types in which the Slicing gesture is embedded.

3.2.1.1 Sequences of different variants of the slicing gesture

In the first example, the Slicing gesture is embedded in a sequence of its different variants. This example is taken from the talk show "Maischberger," where two politicians are discussing their views on the war in Ukraine. The sequence shown in Figure 6 is a response to the host's question, *Ist das so?* ('Is that the case?'). This question addresses the accusation made by the speaker sitting on the left in Figure 6, suggesting that statements from political allies are consistently interpreted as supporting their own political stance.

By saying *Nein* ('No') the speaker shown in the first image in Figure 6 clearly rejects this accusation. He continues by stating that the Finish president has advocated for a certain military strategy in support of Ukraine against Russia. With these verbal utterances the speaker clearly aligns with a significant political viewpoint, reflecting both a shared understanding and an endorsement of collective action. By invoking the Finnish president's advocacy, the speaker bolsters their argument by associating it with a high-profile political figure, thereby not merely stating a fact but emphasizing the necessity and appropriateness of the proposed action of delivering tanks. The inclusion of a direct statement of agreement, *Das ist auch unsere Position* ('This is also our position'), further solidifies the stance by transitioning from merely reporting an opinion to personally endorsing it. This transition is crucial as it positions the speaker within a specific political and ethical framework that advocates for proactive measures in defense matters. The use of the adverb *auch* ('also') and the possessive pronoun *unsere* ('our') implies a collective agreement or policy stance, indicating the speaker's affiliation with the political party to which he belongs, a party that shares this perspective.

The speaker's explanations are co-expressed with a series of Slicing gestures. The first part of his multimodal response features eight two-handed Slicing gestures, with the palms oriented laterally and diagonally within the speaker's gesture space. As shown in the transcript in Figure 6, these manual movements align with specific parts of the speaker's utterance, defining the Finnish president's position and underscoring relevant sections of his statement. The downward movements synchronize with the prosodic prominence in the verbal utterance, specifically with the stress on certain syllables.

The sequence of unstressed and stressed syllables manifests in the upward movement of the hands during the preparational phase and the downward movement during the stroke phase, creating a perceptual rhythm of uniform but emphatic accentuation. This alignment between speech and gesture not only intensifies the semantic content also conveys affective meaning, forming a coherent multimodal emphasis that unfolds temporally.

This rhythmic pattern is interrupted when the speaker explicitly references the position taken by his own party in the debate on delivering tanks to Ukraine. Shifting from using both hands to one hand and changing the orientation of the palm to face his upper body, the speaker states, *Das ist auch unsere Position* ('This is also our position'). This shift in focus is accompanied by three parameters: handedness, palm orientation, and movement pattern, as the speaker executes an arc-like movement directed toward his body. Additionally,

2 Quantitative analyses of this aspect are currently in progress.

FIGURE 6
Example of a gesture sequence embedding different variants of the Slicing gesture. Stroke phases are marked bold, post-stroke holds are underlined, and preparation phases show no formatting. The images are sourced from an interview on the German TV show "Maischberger," aired on January 18, 2023. Link to the interview: https://www.youtube.com/watch?v=8BdJY4RGfuk.

the speaker does not repeat the gestural movement but executes a single gestural stroke, as shown in the final image of Figure 6. These kinesic features make the gesture stand out from the previously created perceptual gestalt, embodying the attentional and discursive shift from the Finnish president to the position of the speaker's party including the speaker himself.

This observation aligns with Harrison's findings on wrist rotation reflecting shifts at the discourse level, such as the "Palm Up to Horizontal Palm Gesture Sequence" (Harrison, 2018, pp. 108–114). The observations presented here build on this research, suggesting that a shift in hand orientation indicates a shift in the speaker's focus within the discourse, particularly towards supporting their own position. The latter complements Calbris (2003, 2011) thorough analysis of the Cutting gesture in French, suggesting that movements along the sagittal axis represent ideas or entities on an interpersonal level, with the speaker committing to one of these ideas. When the gesture moves further away from the speaker's body, it signifies the ideas of another party who may be present.[3] Conversely, if the gesture is close to the speaker's body, it represents ideas or views with which the speaker agrees. This personal character is established by an.

> experiential link (…) between the axis of walking and personal progression or action: it is on this axis that the actor responsible for the action (…), or who is impeded in carrying out his action (…) is situated. This link is itself deduced and confirmed by the confrontation with other examples that include the gestures (…) described here and other gestures characterized by a movement forwards (Calbris, 2003, p. 40).

In the analysis presented, 32 instances of the Slicing gesture directed toward the speaker's body were observed in conjunction with the use of the first-person pronoun. These gestures occurred when speakers were formulating their own position or the position of a group to which they belong, such as a political party or professional group.

3.2.1.2 Sequences of different recurrent gestures

In the example, illustrated in Figure 7, the speaker discusses the public opinion regarding the book he has co-authored, elaborating a critical stance towards the media in Germany. The verbal utterances provide a nuanced example of stance taking, characterized by both affective and evaluative dimensions. These utterances are co-expressed with a sequence of recurrent gestures, within which the Slicing gesture is embedded.

The gesture sequence in question occurs during a discussion about the thesis of a book co-authored by the speaker, as shown in Figure 7. The book critically examines media coverage of the so-called refugee crisis and the war of aggression against Ukraine in Germany. It starts with a two-handed Palm up gesture (Kendon, 2004; Müller, 2004) which is performed with a large movement (first image in Figure 7). It is co-expressed with the rhetorical question *Wo kommt dieser unfassbare Affekt her?* ('Where does this unbelievable affect come from?'), expressing surprise and critique about the intensity of the emotion towards the media reflection published in his book. The speaker's opening question sets the stage for his stance by expressing bewilderment and inviting reflection on the source of a profound emotional response. The use of the adjective *unfassbar* ('unbelievable') intensifies the affective stance, indicating that the speaker finds the level of emotion not only noteworthy but also excessive and potentially problematic. By questioning the origin of this affect, the speaker implicitly positions himself as someone who finds this emotional response unusual and worthy of critical examination.

The both-handed Palm up gesture accompanying the rhetorical question is an instance of the "palm-up epistemic" variant (Cooperrider et al., 2018), described by many authors to convey a consistent range of epistemic meanings (e.g., Chu et al., 2014; Debras, 2017; Mittelberg, 2017; Marrese et al., 2021). This gesture has been associated with epistemic stances related to the absence of knowledge and the subtypes derived from this origin, including uncertainty, obviousness, and interrogatives. The latter is the case observed here, where the gesture accompanies a rhetorical question (Kendon, 2004).

3 This observation is reinforced by the quantitative analysis, which shows that defining the position of others, such as a political party or journalists, is more prevalent when the Slicing gesture is performed with the palm facing the speaker's body. This variant is often moved along the sagittal plane, as Calbris also observed.

FIGURE 7
Example of a gesture sequence embedding different recurrent gestures. Stroke phases are marked bold, post-stroke holds are underlined, pre-stroke holds are marked with a dotted underline, and preparation phases show no formatting. The images are sourced from the German TV show "Markus Lanz," aired on September 29, 2022. Link to the show: https://www.zdf.de/gesellschaft/markus-lanz/markus-lanz-vom-29-september-2022-100.html.

Furthermore, the large movement with which the stroke is performed, characterized by the opening of the hands and a downward motion, embodies the affective meaning of bewilderment and emphasizes the speaker's incredulity.

In the following utterance the speaker turns to the position advocated in his co-authored book: *Wir sagen die ganze Zeit, wir finden dieses Mediensystem großartig. Wir finden die Medienlandschaft in der Bundesrepublik großartig* ('We say all the time that we think this media system is great. We think the media landscape in Germany is great', Figure 7). Here, the speaker adopts an evaluative stance by repeatedly asserting the excellence of the media system and landscape in Germany (*großartig*, 'great'). The repetition of *wir finden* ('we find') emphasizes a joint endorsement by the speaker and his co-author, who is also present in the talk show, suggesting that their view is shared within their collaborative perspective. This collective stance not only strengthens the evaluation but also serves to justify their book against critique, which suggests that the authors criticize the work of the media. By presenting their positive viewpoint as a consistently held and well-established position, the speaker counters the critique leveled against their book. What is more by stating that the media is crucial for democratic functioning, the speaker elevates the significance of their previous positive evaluation thus positioning the speaker firmly in support of the current media landscape.

Notably, the shift in the discursive object from the public's perspective to the authors' stance is mirrored by a corresponding shift in the gestures. The first utterance, *Wir sagen die ganze Zeit, wir finden dieses Mediensystem großartig. Sie ist wichtig für die Demokratie* ('We say all the time that we think this media system is great. It is important for democracy'), is accompanied by a hybrid form of the recurrent Palm up open hand gesture and the Slicing gesture, as illustrated in the second image in Figure 7 (see also Table 2). This "hybrid gesture" (Morris, 2002, p. 58) merges the core meanings of both gestures: defining the authors' position while presenting it on a more or less open hand.

The subsequent utterance, *Wir finden die Medienlandschaft in der Bundesrepublik großartig. Sie ist wichtig für die Demokratie* ('We think the media landscape in Germany is great. It is important for democracy'), is accompanied by multiple strokes of the Slicing gesture, shown in the third image of Figure 7. These gestures not only establish the media system as a discursive object but also emphasize the authors' positive stance toward it.

The following limitation of this positive evaluation is accompanied by the Stretched index finger, a gesture described as expressing attention (Müller et al., 2013; Bressem and Müller, 2014). This gesture marks the contrasting stance established in speech, signaling a shift from a general praise to a more nuanced position. By acknowledging 'tendencies in development' (*Tendenzen in der Entwicklung*) that require attention, the speaker adopts a critical stance. This critical stance does not negate the earlier positive evaluations but adds complexity to the speaker's position, indicating awareness of and concern for ongoing changes that might threaten the media's positive role. This more nuanced and critical position is embodied by the Ring gesture (Müller, 2014b) the speaker uses twice, which conveys the idea of making a precise point.

The speaker's stance in these utterances is multifaceted, combining strong positive evaluations with a cautious critical perspective. The single gestures used in this gestural sequence are highly coordinated and aligned with ideas expressed in speech. The two-handed Palm up gesture expresses bewilderment and critique, while the hybrid Palm up open hand and Slicing gesture emphasizes positive evaluations. The transition to a critical stance is marked by the Attention gesture (Stretched index finger), and the Ring gesture underscores specific concerns, effectively mirroring the progression of the speaker's verbal message. This integrated use of gestures and speech demonstrates the nuanced expression of complex stances in political discourse.

3.2.1.3 Sequences of different gesture types

The following example illustrates the embedding of the Slicing gesture within a sequence of different gesture types including singular gestures (also known as depictive gestures or iconic gestures). In the example illustrated in Figure 7, the speaker juxtaposes different interpretations of the ban on violence in international law, arguing that the West bends the law to fit its own goals when necessary. As he begins to elaborate on his argument, he uses the Slicing gesture, which is embedded in a sequence of three gestures. Notably, each idea in his utterance is accompanied by a corresponding gesture.

The beginning of his utterance *Aber mein übergreifendes Argument dabei* ist ('But my overarching argument here is') is accompanied by the Slicing gesture. While producing the possessive pronoun *mein* ('my') the two-handed version of the Slicing gesture is performed. The hands are moved down while being placed in the central gesture space. The gesture essentially cuts out and isolates an idea (in this case, an argument that is verbally referenced later) and defines it. The gesture is followed by a manual movement showing a pincer-like configuration, where both the thumb and the index finger are spread apart and slightly bent (second image in Figure 8). The gesture is moved in an arc to the right of the speaker while he says *übergreifend* ('overarching'; literally: 'over-grasping'). The gesture basically embodies both morphemes of the adjective: *über* ('over') is depicted by the arced movement to the right as if the gesture spans several entities, *greifend* ('grasping') is embodied by the pincer-like configuration of the hand. The gesture involves a fist being repeatedly moved downward while the speaker says *Argument ist dabei* ('argument here is', see bold characters in the third image in in Figure 8 for the position of the strokes). This gesture depicts the holding of an object while emphasizing its strength as well as the speaker's involvement in the discussion with the downward movement.

In this example, the Slicing gesture brings the speaker's view on the handling of the international law into the focus of attention. By being simultaneously expressed with the possessive pronoun *mein* ('my'), it essentially isolates (or cuts out) his argument from others and clarifies its definition. The subsequent gesture alludes to this argument being isolated and further elaborates on it both verbally and gesturally. The third gesture physically grasps the isolated argument in the hand, emphasizing it with a repetitive back-and-forth movement.

This example differs not only in the structural components, i.e., the gesture types, but also in its temporal dynamic. This sequence is characterized by a quick succession of gestures leading to the fact that almost each word is accompanied by a new gesture.

3.2.2 Developing meta-pragmatic meaning in gesture sequences of slicing gestures

The examples presented in the previous sections show a tight interplay of gesture and speech in the meaning-making process. This means that the gestures discussed acted on individual parts of speech and form multimodal units with them. However, we also observed cases, in which the Slicing gesture became more detached from speech and thus appeared to have developed a second-order meaning. This phenomenon was particularly evident in cases of extended gesture sequences in which the gestural strokes exceeded five repetitions. An example is given in Figure 9.

The whole multimodal explanation lasts 24 s. Its temporal unfolding is visualized by the arrow in Figure 9. The single gestural strokes are placed on the arrow by means of vertical lines. Their accompanying speech units are set above it. While answering the question addressed by the host, the speaker uses 22 Slicing gestures strokes. In the majority of cases (16 gestures), the speaker uses a single hand with the palm oriented diagonally toward the center (PLdiTC, see also Figure 2B). In contrast, six gestures feature the palm facing the speaker's body (PLTB, see also Figure 2A). In one case, the two-handed version is used which appears to be the only gesture in this sequence which aligns thematically with speech. To be more precise, in this instance, the Slicing gesture defines a moment in time when the speaker says *jedenfalls im moment* ('at least for the moment'). The hands are moved down co-expressively with the adverb *jedenfalls* ('at least') and held while saying *im moment* ('at the moment'). The other strokes of the Slicing gesture are aligned temporally with speech, but they do not operate on single discursive objects verbalized. This means they do not primarily define a moment in time or a discourse object. These versions of the Slicing gesture are employed by the speaker to consistently integrate his multimodal utterances thematically [catchment in McNeill's (1992)], signaling their cohesive involvement in a single communicative activity. What's more, they

FIGURE 8
Example of a gesture sequence embedding different gesture types including the Slicing gesture. Stroke phases are marked bold, preparation phases show no formatting. The images are sourced from the German TV show "Unter den Linden," aired on September 19, 2022. Link to the show: https://www.ardmediathek.de/video/unter-den-linden/wendepunkt-erodiert-putins-macht-chance-auf-frieden/phoenix/Y3JpZDovL3Bob2VuaXguZGUvMjkzMjU0OA.

FIGURE 9
Example of an extended Slicing gesture sequence with metapragmatic meaning. Stroke phases are marked bold, post-stroke holds are underlined, pre-stroke holds are marked with a dotted underline, and preparation phases show no formatting. The example is sourced from the German TV show "Unter den Linden," aired on September 19, 2022. https://www.ardmediathek.de/video/unter-den-linden/wendepunkt-erodiert-putins-macht-chance-auf-frieden/phoenix/Y3JpZDovL3Bob2VuaXguZGUvMjkzMjU0OA.

also visually convey the speaker's intention to clearly define one of the views discussed in the talk show. The speaker's alignment with this view is evident not only through his verbal expression but also through numerous Slicing gestures directed towards his body, reflecting an interpersonal level of argumentation.

The Slicing gestures in this example demonstrate a higher level of independence from individual speech components, operating on a meta-level. Rather than anchoring themselves to specific propositions or verbalized elements of the speaker's discourse, the frequent use of this gesture creates a dynamic characterized by a broader rhetorical stance—one of clarity and sharpness, effectively 'defining things clearly.' As an embodied discursive strategy that characterizes communicative behavior as a social act (Lempert, 2011, p. 258), this metapragmatic meaning becomes particularly evident in sequences of the Slicing gesture. Thus, the analysis shifts from isolated speech or gesture elements to more extended, cohesive patterns of gesture use. These patterns, or longer sequences, signal a higher level of communicative intent, positioning the Slicing gesture as a powerful tool for shaping meaning and interaction beyond the immediate verbal content. This aligns with Lempert's (2011) argument, highlighting the importance of considering larger units of analysis in understanding communicative behavior.

In essence, the Slicing gesture frequently occurs in sequences featuring various recurrent gestures or different variants of the Slicing gesture. In the former scenarios, this gesture is closely connected to the discursive objects verbalized in the utterances. When employed in extended sequences of Slicing gestures, it embodies the metapragmatic meaning of defining things with clarity and sharpness.

4 Kinesthetic experiences as fundamental to the meaning of the slicing gesture

The quantitative analysis of the Slicing gesture reveals its prevalent use in defining discourse objects and conveying metapragmatic meanings of sharpness and decisiveness. These functions are frequently integrated into stance-taking activities, including evaluative, deontic, and affective stances. The gesture's suitability for these communicative purposes stems from its physical characteristics, which embody a "categorical or decisive character" (Calbris, 2011, p. 116). Specifically, the downward movement, the flat hand, which exposes edge of the hand, work together to symbolize a slicing action. This movement effectively demarcates concepts or delineates positions, thereby reinforcing the speaker's decisiveness and clarity in their stance. It emphasizes a sense of finality or closure, which aligns with the decisive and categorical nature of the Slicing gesture.

> The categorical aspect (…) stems from the perceptual schema of cutting, of separating: 'that is it; that is not it'. By analogy with the cutting edge of a blade or an axe, lowering the edge of the hand relates back to cutting, physically and psychologically 'slicing', i.e., to deciding, to stopping something, to being categorical. By extension, the gesture evokes the categorical character of an assertion, a fact or a principle (Calbris, 2003). Thus, regardless of the configuration, the lowering of the edges of rigid, vertically held hands represents the action of slicing, separating, stopping and, hence, the act of deciding (Calbris, 2011, p. 121).

The act of deciding, Calbris refers to, can literally be seen as embodied by this gesture given the etymology of the word "decisive" whose stem is derived from the Latin word *decidere* which means "to cut off" (in the sense of decide)[4]. What is more, the Slicing gesture's capacity in communicating sharpness and decisiveness is intimately linked with the quality of movement and thus with the sensorimotor experience of this gesture. Accordingly, signification resides in the gesture itself and the physical sensation of performing it. The speaker experiences these proprioceptive qualities firsthand, reinforcing their sense of actively defining and categorizing concepts. This embodied experience allows them to internalize the sensation of sharpness, decisiveness and authority. Consequently, the gesture not only communicates these aspects to the interlocutor but also reinforces the

4 Online Etymology Dictionary, s.v. decisive (https://www.etymonline.com/search?q=decisive).

speaker's own self-perception and stance, contributing to a second-order meaning of being sharp and decisive in their communication.

The proprioceptive dimension referred to here, also plays a crucial role in understanding the meaning of this gesture. As Calbris (2003) emphasized regarding the Cutting gesture in French, the interlocutors empathically perceive the proprioceptive qualities of the gesture, such as muscular tension and control. The abrupt nature of an individual's Slicing gestures induces similar sensations of rigidity and sudden motion in the interlocutor's body. The characteristics of power, control, and sharpness signaling determination, boundaries, or emphasis within the context of discourse are not only felt by the speaker but also by the interlocutor, making understanding a gesture a not only a visual but physical experience.[5]

This dimension of meaning-making has often been overlooked in gesture studies that prioritize gestures' visuo-spatial imagery. In this traditional view, gestures represent actions or objects through their iconicity, with production and understanding depending primarily on visual properties. However, the idea that gestures convey meaning through the manual actions they are based on (e.g., Kendon, 2004; Müller, 2004; Streeck, 2009) already challenges the dominance of imagery in the meaning-making process. Recent research increasingly highlights the proprioceptive dimension of gestures and other body movements (e.g., Müller and Ladewig, 2013; Streeck, 2017; Müller and Kappelhoff, 2018; Ladewig and Hotze, 2021), emphasizing its central role in the meaning-making process of gestures. Accordingly, the meaning of being sharp, decisive and clear not only relies on visual qualities of the gesture but is particularly related to the "embodied-affective aspects" (Gallagher, 2017, p. 152) of the configuration and movement of the hand. "The uniqueness of the dynamics is first and foremost a kinesthetic uniqueness, not a visual uniqueness" (Sheets-Johnstone, 2003, p. 74). Therefore, the body is never absent or transparent but contributes to the speaker's and interlocutors experience of a situation (see Colombetti, 2014). Movements are felt and can be controlled, and they are understood through the body (ibid., see Kappelhoff and Müller, 2011; Horst et al., 2014; Müller and Kappelhoff, 2018; Ladewig and Hotze, 2021, for phenomenological approaches to gestures that have evolved in recent years). The kinesthetic experiences evoked by the gestural form are fundamental to its meaning and help stabilize its interpretation (see Sheets-Johnstone, 2003). The gesture's proprioceptive qualities—such as power, control, tension, and sharpness—play a crucial role in its communicative function. The hand, acting as a manual blade, performs a cutting motion that isolates and highlights specific concepts and viewpoints. This act of delineation makes the gesture an exceptionally potent semiotic resource for expressing a stance. By embodying the physical sensations associated with decisiveness, sharpness and clarity, the gesture effectively conveys the speaker's commitment to a discussion, making their position clear to the audience.

5 One possible explanation for how interlocutors may experience a proprioceptive sense of a seen gesture lies in the concept of mirror neurons. These neurons are thought to allow observers to internally simulate observed actions, which might account for how the Slicing gesture's qualities of power, control, and sharpness are "felt" by the interlocutor. For a broader discussion, see Gallese et al. (1996) on mirror neuron systems and their role in embodied communication.

5 Conclusion

This paper investigated the use of the Slicing gesture in mediated political communication, particularly focusing on German televised political talk shows, with a primary emphasis on stance-taking. The Slicing gesture is identified as a recurrent gesture frequently used to define discourse objects and convey sharpness and decisiveness. These meanings are embodied by the gesture's physical characteristics, such as its downward movement and flat hand configuration, which symbolize a slicing action that demarcates concepts and positions, reinforcing the speaker's decisiveness and clarity.

Quantitative analyses reveal that the Slicing gesture is embedded in the speech activity of stance-taking, encompassing evaluative, deontic, and affective stances. In expressing these versatile meanings of stance, the Slicing gesture is often integrated into gesture sequences, which can include different variants of the Slicing gesture, various recurrent gestures, or different gesture types. These sequences are particularly prevalent in mediated contexts, such as political talk shows, where longer gesture sequences are more common than in private settings. Also extended sequences of variants of the Slicing gesture are common in televised political communication. In these cases, the gesture appears to develop momentum, no longer operating on the level of single parts of speech but adding a metapragmatic dimension to the discourse. The metapragmatic meaning 'defining things with clarity and sharpness' is diagrammatically motivated by the juxtaposed Slicing gestures, creating a perceptual gestalt that reinforces the speaker's engagement and stance in the debate demonstrating that taking stance is "an essentially *interactive activity*" (Kärkkäinen, 2003, p. 183, emphasis in the original). What is more, the speaker present themselves in this way as someone who clearly names things and positions themselves in the discussion by differentiating from their fellow discussants, thereby taking an interpersonal stance.

The study argued that the meaning-making process of the Slicing gesture, along with its effectiveness in communicating clarity when making arguments, extends beyond its visual properties, emphasizing the importance of its proprioceptive dimension. This dimension involves the physical sensations associated with performing the gesture, such as muscular tension, control power, and sharpness. The speaker experiences these proprioceptive qualities firsthand, while interlocutors empathically perceive them, making the understanding of a gesture not only a visual experience but also a physical one.

The kinesthetic sensations of sharpness, determination, and control which are qualities integral to the Slicing gesture, make it an effective semiotic resource for stance-taking and for "self-presentation" (Streeck, 2008) in political communication. These proprioceptive qualities not only reinforce the speaker's self-perception of decisiveness and determination but also communicate these attributes to the audience. This alignment between physical sensation and communicative intent forms the basis for the habitualization and stabilization of this gesture. "(H)ow stances are taken, and which stances are taken, are often habitually repeated by people with similar identities" (Kiesling, 2022, p. 412). Thus, it is unsurprising that the Slicing gesture emerges as a prevalent gesture in the domain of political communication. "[G]estures reveal a great deal about interactional practices, the social norms that underlie them, and how local and wider ideologies in societies shape the nature of gestures and their use" (Brookes and Le Guen, 2019, p. 129). Accordingly, if certain practices, norms, and ideologies are shared among speech communities, they may share gestural forms.

Data availability statement

The dataset analyzed in this study is not publicly available; however, the data used in the analysis are included within the article. Questions regarding further information on this data should be directed to the corresponding author.

Author contributions

SL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Support was granted by the Open Access Publication Funds of Göttingen University.

Acknowledgments

I would like to thank Mathias Roloff for his detailed work on the gesture drawings. My sincere thanks to the reviewers for their efforts and valuable suggestions.

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

Agha, A. (2006). Language and social relations. Cambridge: Cambridge University Press.

Andries, F., Meissl, K., de, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). Multimodal stance-taking in interaction—a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Bohmann, A., and Ahlers, W. (2022). Stance in narration: finding structure in complex sociolinguistic variation. *J. Socioling.* 26, 65–83. doi: 10.1111/josl.12533

Bressem, J. (2013). "A linguistic perspective on the notation of form features in gestures" in Body – Language – Communication: An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke. S. H. Ladewig, D. McNeill and S. Teßendorf (Berlin, New York: Mouton de Gruyter), 1079–1098.

Bressem, J. (2021). Repetitions in gesture: A cognitive-linguistic and usage-based perspective. Berlin: De Gruyter.

Bressem, J., Ladewig, S. H., and Müller, C. (2013). "A linguistic annotation system for gestures (LASG)" in Body – Language – Communication: An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Teßendorf (Berlin, Boston: De Gruyter Mouton), 1098–1125.

Bressem, J., and Müller, C. (2014). "A repertoire of German recurrent gestures with pragmatic functions" in Body – Language – Communication. An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin/Boston: De Gruyter Mouton), 1575–1592.

Bressem, J., and Müller, C. (2017). The "negative-assessment-construction"–a multimodal pattern based on a recurrent gesture? *Linguist. Vanguard* 3:20160053. doi: 10.1515/lingvan-2016-0053

Brookes, H., and Le Guen, O. (2019). Gesture studies and anthropological perspectives. *Gesture* 18, 119–141. doi: 10.1075/gest.00040.bro

Calbris, G. (2003). From cutting an object to a clear cut analysis. *Gesture* 3, 19–46. doi: 10.1075/gest.3.1.03cal

Calbris, G. (2011). Elements of meaning in gesture, vol. *5*. Amsterdam: John Benjamins Publishing Company.

Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *J. Exp. Psychol. Gen.* 143, 694–709. doi: 10.1037/a0033861

Colombetti, G. (2014). The feeling body: Affective science meets the enactive mind. Cambridge, Massachusetts: MIT press.

Cooperrider, K., Abner, N., and Goldin-Meadow, S. (2018). The palm-up puzzle: meanings and origins of a widespread form in gesture and sign. *Communication* 3. doi: 10.3389/fcomm.2018.00023

Debras, C. (2017). The shrug. *Gesture* 16, 1–34. doi: 10.1075/gest.16.1.01deb

du, J. (2007). The stance triangle. *Stancetaking Discourse* 164, 139–182. doi: 10.1075/pbns.164.07du

Du Bois, J. W., and Kärkkäinen, E. (2012). Taking a stance on emotion: affect, sequence, and intersubjectivity in dialogic interaction. *Text Talk* 32, 433–451.

Englebretson, R. (2007). Stancetaking in discourse: an introduction. *Stancetaking Discourse* 164, 1–25. doi: 10.1075/pbns.164.02eng

Fricke, E. (2010). Phonaestheme, Kinaestheme und multimodale Grammatik: Wie Artikulationen zu Typen werden, die bedeuten können. *Sprache Literatur* 41, 69–88. doi: 10.1163/25890859-041-01-90000005

Gallagher, S. (2017). Enactivist interventions: Rethinking the mind. Oxford: Oxford University Press.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain J. Neurol.* 119, 593–609. doi: 10.1093/brain/119.2.593

Goffman, E. (1956). The presentation of self in everyday life. Edinburgh: University of Edinburgh, Social Sciences Research Centre.

Harrison, S. (2018). The impulse to gesture: Where language, minds, and bodies intersect. Cambridge: Cambridge University Press.

Harrison, S., and Ladewig, S. H. (2021). Recurrent gestures throughout bodies, languages, and cultural practices. *Gesture* 20, 153–179. doi: 10.1075/gest.21014.har

Horst, D., Boll, F., Schmitt, C., and Müller, C. (2014). "Gestures as interactive expressive movement: inter-affectivity in face-to-face communication" in Body – Language – Communication: An international handbook on multimodality in human interaction. eds. A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin, Boston: De Gruyter Mouton), 2112–2125.

Jakobson, R. (1960). "Closing statement. Linguistics and poetics" in Style in language. ed. T. A. Sebeok (Cambridge, MA: MIT Press), 350–449.

Johnstone, B. (2017). "Characterological figures and expressive style in the enregisterment of linguistic variety" *Language and a sense of place: studies in language and region*, 283–300.

Kappelhoff, H., and Müller, C. (2011). Embodied meaning construction. *MSW* 1, 121–153. doi: 10.1075/msw.1.2.02kap

Kärkkäinen, E. (2003). Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think. Amsterdam: John Benjamins Publishing Company.

Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk* 26, 699–731. doi: 10.1515/TEXT.2006.029

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *J. Pragmat.* 23, 247–279. doi: 10.1016/0378-2166(94)00037-F

Kendon, A. (2004). Gesture. Visible action as utterance. Cambridge: Cambridge University Press.

Kiesling, S. F. (2022). Stance and stancetaking. *Ann. Rev. Linguist.* 8, 409–426. doi: 10.1146/annurev-linguistics-031120-121256

Kita, S. (2000). "How representational gestures help speaking" in Language and gesture. ed. D. McNeill (Cambridge: Cambridge University Press), 162–185.

Kok, K. (2016). The status of gesture in cognitive-functional models of grammar. Utrecht: LOT Publications.

Ladewig, S. H. (2014a). "The cyclic gesture" in Body – Language – Communication. An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin, Boston: De Gruyter Mouton), 1605–1618.

Ladewig, S. H. (2014b). "Recurrent gestures" in Body – Language – Communication. An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin, Boston: De Gruyter Mouton), 1558–1575.

Ladewig, S. H. (2024). "Recurrent gestures: cultural, individual, and linguistic dimensions of meaning-making" in The Cambridge handbook of gesture studies. ed. A. Cienki (Cambridge: Cambridge University Press), 32–55.

Ladewig, S. H., and Hotze, L. (2021). The slapping movement as an embodied practice of dislike: inter-affectivity in interactions among children. *Gesture* 20, 285–312. doi: 10.1075/gest.21013.lad

Lempert, M. (2008). The poetics of stance: text-metricality, epistemicity, interaction. *Lang. Soc.* 37, 569–592. doi: 10.1017/S0047404508080779

Lempert, M. (2011). Barack obama, being sharp. *Gesture* 11, 241–270. doi: 10.1075/gest.11.3.01lem

Lempert, M. (2017). Uncommon resemblance: pragmatic affinity in political gesture. *Gesture* 16, 35–67. doi: 10.1075/gest.16.1.02lem

Marrese, O. H., Raymond, C. W., Fox, B. A., Ford, C. E., and Pielke, M. (2021). The grammar of obviousness: the palm-up gesture in argument sequences. *Front. Commun* 6:663067. doi: 10.3389/fcomm.2021.663067

McNeill, D. (1992). Hand and mind. What gestures reveal about thought. Chicago: University of Chicago Press.

McNeill, D. (2005). Gesture and thought. Chicago: University of Chicago Press.

Mittelberg, I. (2017). Multimodal existential constructions in German: manual actions of giving as experiential substrate for grammatical and gestural patterns. *Linguistics Vanguard* 3, 1–14. doi: 10.1515/lingvan-2016-0047

Morris, D. (1994). Bodytalk: The meaning of human gesture. New York: Crown Trade.

Morris, D. (2002). Peoplewatching. London: Random House.

Mosher, J. A. (1916). The essentials of effective gesture for students of public speaking. Norwood Massachusetts: Norwood Press.

Müller, C. (1998). Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich. Berlin: Arno Spitz.

Müller, C. (2004). "Forms and uses of the palm up open hand. A case of a gesture family?" in Semantics and pragmatics of everyday gestures. eds. C. Müller and R. Posner (Berlin: Weidler), 234–256.

Müller, C. (2010). Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. *Sprache Literatur* 41, 37–68. doi: 10.1163/25890859-041-01-90000004

Müller, C. (2014a). "The ring across space and time: variation and stability of forms and meanings" in Body – Language – Communication: An international handbook on multimodality in human interaction. Handbooks of linguistics and communication science. eds. A. C. Cornelia Müller, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin/Boston: De Gruyter Mouton), 1511–1522.

Müller, C. (2014b). "Ring-gestures across cultures and times: dimensions of variation" in Body – language – communication. An international handbook on multimodality in human interaction. eds. A. Cornelia Müller, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin/Boston: De Gruyter Mouton).

Müller, C. (2017). How recurrent gestures mean. *Gesture* 16, 277–304. doi: 10.1075/gest.16.2.05mul

Müller, C., Bressem, J., and Ladewig, S. H. (2013). "Towards a grammar of gesture: a form-based view" in Body – Language – Communication. An international handbook on multimodality in interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and S. Tessendorf (Berlin, Boston: De Gruyter Mouton), 707–733.

Müller, C., and Kappelhoff, H. (2018). Cinematic metaphor: Experience–affectivity–temporality. Berlin/Boston: Walter de Gruyter.

Müller, C., and Ladewig, S. H. (2013). "Metaphors for sensorimotor experiences. Gestures as embodied and dynamic conceptualizations of balance in dance lessons" in Language and the creative mind. eds. B. Dancygier, J. Hinnell and M. Borkrent (Stanford: CSLI), 295–324.

Neumann, R. (2004). "The conventionalization of the ring gesture in German discourse" in The semantics and pragmatics of everyday gestures. eds. C. Müller and R. Posner (Berlin: Weidler), 217–223.

Ott, E. A. (1902). How to gesture. New York City: Hinds & Noble.

Quintilian, M. F. (1969). The Institutio Oratoria of Quintilian with an English translation by H. E. Butler. The Loeb classical library. New York: G. P. Putnam and Sons.

Seyfeddinipur, M. (2004). "Meta-discursive gestures from Iran: some uses of the 'Pistol Hand'" in The semantics and pragmatics of everyday gestures. eds. C. Müller and R. Posner (Berlin: Weidler), 205–216.

Sheets-Johnstone, M. (2003). Kinesthetic memory. *Theoria et Historia Scientiarum*. 7, 69–92. doi: 10.12775/ths.2003.005

Silverstein, M. (2004). "Cultural" concepts and the language-culture nexus. *Curr. Anthropol.* 45, 621–652. doi: 10.1086/423971

Streeck, J. (2008). Gesture in political communication: a case study of the democratic presidential candidates during the 2004 primary campaign. *Res. Lang. Soc. Interact.* 41, 154–186. doi: 10.1080/08351810802028662

Streeck, J. (2009). Gesturecraft: Manufacturing understanding. Amsterdam / Philadelphia: John Benjamins.

Streeck, J. (2017). Self-making man: A day of action, life, and language. New York, NY, USA: Cambridge University Press.

Stubbs, M. (1996). Towards a modal grammar of English: A matter of prolonged fieldwork. Blackwell, Oxford: Text and Corpus Analysis.

Teßendorf, S. (2014). "Pragmatic and metaphoric gestures– combining functional with cognitive approaches" in Body – Language – Communication. An international handbook on multimodality in human interaction. eds. C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill and J. Bressem (Berlin / Boston: De Gruyter Mouton), 1540–1558.

Tutton, M. (2011). How speakers gesture when encoding location with English on and French Sur. *J. Pragmat.* 43, 3431–3454. doi: 10.1016/j.pragma.2011.07.012

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research", in: Proceedings of 5th International Conference on Language Resources and Evaluation (LREC 2006), 1556–1559.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 11 February 2025
DOI 10.3389/fpsyg.2024.1461123

Check for updates

OPEN ACCESS

EDITED BY
Geert Brône,
KU Leuven, Belgium

REVIEWED BY
Danyal Farsani,
Norwegian University of Science and Technology, Norway
Michael Winter,
Julius Maximilian University of Würzburg, Germany

*CORRESPONDENCE
Barbara Laner
barbara.laner@germanistik.uni-freiburg.de

RECEIVED 07 July 2024
ACCEPTED 02 December 2024
PUBLISHED 11 February 2025

CITATION
Laner B (2025) Mobile stance-taking in nature: an exploration of gaze patterns during assessments of objects in nature.
Front. Psychol. 15:1461123.
doi: 10.3389/fpsyg.2024.1461123

COPYRIGHT
© 2025 Laner. 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.

Mobile stance-taking in nature: an exploration of gaze patterns during assessments of objects in nature

Barbara Laner*

University of Freiburg, Freiburg im Breisgau, Germany

In this paper, I examine the interactional dynamics of walkers assessing entities in nature, focusing on gaze behavior during these sequences. The analysis is based on a corpus of 10 hiking pairs who walked through the Black Forest National Park while wearing mobile eye-tracking glasses to record their gaze behavior and verbal practices. Using a combined quantitative and qualitative approach, the research identifies gaze patterns in 127 sequences and highlights the role of bodily-visual practices. Contrary to existing literature, the findings indicate that mutual gaze in this setting is not used to mark affiliation but instead occurs only during strong disagreements about initial assessments. During agreements, walkers maintain a triangular position, both gazing at the assessable object without looking at each other. Thus, in this context, gazing at each other serves different interactional purposes, as this study will demonstrate.

KEYWORDS

assessments, stance-taking, affective stances, gaze, preference organization, walking and talking, mobile interactions, noticing

1 Introduction

Expressing viewpoints and evaluating objects is omnipresent in everyday interaction. Whether we discuss food preferences, exchanging political views, or talking about our favorite movies, we are constantly expressing opinions and evaluating the world around us (*cf.* among others Siromaa and Rauniomaa, 2021: 96). So, it is also common to share stances toward entities in the surrounding environment while experiencing nature together, such as during a walk through the forest (*cf.* Auer et al., 2024; Botsch et al., 2025). Using gazes, gestures, body positions, movements, and language, different aspects in the changing environment can be made relevant. Interactants produce interactional noticings (*cf.* Sacks, 1995; Schegloff, 2007: 87, FN17) whenever they refer to something in nature and make it relevant for the interaction. In the same turn, speakers frequently express their stances toward the noticed object (e.g., look how beautiful that is). Our mutual understanding of the world around us is thus led by sharing observations and discoveries and negotiating evaluative stances toward these.

Research has shown that gaze behavior plays a crucial role during assessment sequences, as "gaze directions and gaze shifts in particular sequential positions in interaction have an important role, first, in relation to one's right to assess a referent […] and, second, in relation to one's ability to assess a referent" (Haddington, 2006: 283). By establishing or avoiding eye contact, we can direct attention, signal agreement or disagreement, and even express power dynamics (see discussion below, Section 2.3). Therefore, the study of gaze is an important aspect of analyzing stance-taking in human interactions. Although the importance of gaze behavior is undisputed, most studies focus on non-mobile settings (e.g., Kendrick and Holler, 2017; Siromaa and Rauniomaa, 2021;

Pekarek Doehler et al., 2021; Krug, 2025), even though gaze behavior is influenced by moving through space, such as while walking (*cf.* Auer and Laner, 2025).

The activity type I am interested in is joint walks in the forest, where participants need to pay attention to the path in addition to nature and possible assessables, which influences their gaze patterns. I will focus on gaze patterns within sequences in which objects in nature are assessed, i.e., in which evaluative stances are shared. Specifically, I will examine gaze behavior in two parts of the sequence: first, the gaze behavior before and during the first assessment, and second, the gaze behavior when the recipient responds to the assessment. I aim to show that gaze patterns during the first part align with the findings for static interactions described by Haddington (2006, p. 284), but deviate in the second part of the sequence.

In Section 2, I will discuss the theoretical background of the researched activity type, noticings and their cooccurrence with assessments, and findings on gaze behavior during assessment sequences in more detail. Section 3 will provide a methodological overview, before I will present my analysis in Section 4. I will start with a quantitative overview of the data and general findings, followed by a qualitative analysis of gaze patterns during the assessment sequences. The findings will then be discussed in the conclusion in Section 5.

2 Background

2.1 Walking through nature together

Walking together is not only to be understood as a fundamental form of locomotion but also as a social practice. Social science studies, which have been interested in walking together as "doing walking" since the 1970s (Ryave and Schenkein, 1974: 265), describe this as a joint achievement by the participants and as socially and physically co-organized action. Whenever two people walk together, they usually do so in a side-by-side configuration where walking and stopping, talking, and even gaze must be coordinated. This coordination occurs interactively, and various studies on walking together (Mondada, 2017; Merlino and Mondada, 2019) show that this is a very orderly process and constitutes a social practice characterized by strong mutual "monitoring" (*cf.* Deppermann and Schmitt, 2007: 121; Stukenbrock, 2015: 54).

Building on this understanding of walking as a coordinated social practice, walking through a forest adds additional layers of complexity to interaction. In this natural setting, participants not only need to coordinate their movements and conversations but also navigate the terrain and engage with the surrounding environment. The forest provides a dynamic backdrop where the interplay of nature and social interaction becomes particularly evident. As walkers encounter various natural objects and landmarks, they frequently pause to observe, assess, and discuss these elements, further enriching their joint walking experience.

Joint walks through nature usually do not solely aim to transfer participants efficiently from point A to point B, nor do they exclusively serve as occasions for talk without a directional goal. Walks or hikes in nature typically feature intermittent phases of silence and conversation. Even when not engaged in conversation, walkers are in an "open state of talk" (Goffman, 1981: 134). Additionally, conversations during walks also include both displaced and situated speech. Displaced speech has no topical restrictions[1], while situated speech during forest walks often refers to environmental features that contribute to spatial orientation, wayfinding, or the shared experience of nature [c.f. for a more detailed discussion Auer et al. (2024)].

Experiencing nature together is an integral part of forest walks (Rauniomaa et al., 2019; Lehmann, 1999; Burckhardt, 2006; Botsch, 2021; Auer et al., 2024; Botsch et al., 2025) and can even be one of the main purposes of going on a hike through the National Park. As a result, these walks centrally involve acts of referring to and discussing noteworthy objects in the surroundings. I am interested in how walkers use verbal and bodily resources to share and express their stances toward certain aspects or objects in nature, with a particular focus on their gaze behavior during these sequences.

2.2 Noticing and assessing

The social action of making something relevant in the immediate environment (and calling joint attention to it) has been greatly discussed in literature as the action of 'doing a noticing' (*cf.* Auer et al., 2024; Goodwin and Goodwin, 2012; Kääntä, 2014; Keisanen, 2012; Laanesoo and Keevallik, 2017; Laner, 2022; Sacks, 1992, 1995; Schegloff, 1988, 2007; Stivers and Rossano, 2010: 9; Szymanski, 1999). Through an interactional noticing (*cf.* Sacks, 1995; Schegloff, 2007: 87, FN17), an object or phenomenon can be made relevant as noticeable in the immediate environment (Schegloff, 1986; Schegloff and Sacks, 1973). Schegloff (2007: 219) describes this as a "source/outcome" relationship, where the speaker retrospectively reacts to a perceived entity and highlights it as noticeable. From the recipient's perspective, an interactional noticing stands in a sequentially first position (*cf.* also Keisanen, 2012: 201). However, the social action of "calling joint attention to and achieving intersubjectivity over a selected publicly perceivable referent" (Pillet-Shore, 2020: 7) can also include assessments (*cf.* Golato, 2002; Goodwin and Goodwin, 1986; Goodwin and Goodwin, 1987; Heritage, 2002; Heritage and Raymond, 2005; Pomerantz, 1984; Stivers and Rossano, 2010). As an example, assessments can occur in combination with perception imperatives (e.g., 'look how beautiful that tree is', *cf.* also Laner, 2022).

It becomes evident that a clear separation between noticing and assessment is not always possible. Goodwin and Goodwin (2012: 275) also note that an assessment can simultaneously be a noticing. In this paper I will focus on sequences in which assessments occur in sequentially first positions and thereby mostly also function as noticings.

2.3 Gaze behavior during assessments

Gaze behavior is a fundamental aspect of human interaction that has garnered considerable attention in interactional linguistics. Its role in social dynamics, particularly during evaluative stance-taking sequences, is multifaceted and pivotal. As asserted by Goodwin (1981), gaze itself can become socially relevant, transforming into a

1 *cf.* Auer and Laner (2025) on laughables that are produced during displaced speech in forest walks.

crucial tool for communication. This notion finds further support in the work of Haddington (2006: 299), who argues that mutual gaze expresses convergent stances, underscoring its importance in signaling agreement or alignment during interactions involving evaluation. Agreeing second assessments are preferred, while disagreeing second assessments are dispreferred in most contexts. In more detail, he elucidates distinct functions of gaze in the context of assessments. Firstly, gaze acts as a means for interactants to identify and focus on assessable objects, facilitating the construction of a shared *participation framework* (Goodwin and Goodwin, 2004). This use suggests a visual grounding mechanism that reinforces joint attention and participation in assessing activities. Secondly, it is further argued that mutual gaze between two interlocutors often accompanies their expressions of agreement regarding an assessable, highlighting gaze as a tool for displaying mutual understanding and like-mindedness, reinforcing social solidarity. Thirdly, speakers integrate gaze with verbal and non-verbal cues to position themselves in relation to the stances proposed by their coparticipants, either affiliating with or diverging from these positions. Finally, listeners interpret coparticipants' gaze trajectories as cues to understand speakers' stances, contributing to the ongoing negotiation of meaning and interactional alignment (*cf.* for more detail Haddington, 2006).

Kendrick and Holler (2017: 2) shed further light on the expressive function of gaze, noting its role in moderating the level of arousal and emotionality. They propose that gaze direction in the context of polar questions serves as a resource for constructing affiliative and disaffiliative actions, which aligns with Kidwell's (2006) and Haddington's (2006) findings, emphasizing its significance in stance-taking and maintaining social solidarity. The research by Kendrick and Holler (2017) illuminates how gaze patterns contextualize responses, with preferred answers (agreeing second assessments) often accompanied by mutual gaze and dispreferred ones (disagreeing second assessments) met with averted gazes. This dichotomy underscores the nuanced role of gaze in managing conversational preferences and social dynamics. Similar, interactional data from various languages reveals a recurrent multimodal practice that respondents deploy in turn-initial positions in dispreferred responses to actions such as information requests, assessments, proposals, and informing (see Robinson, 2020): This practice involves the verbal delivery of expressions equivalent to the English 'I do not know' and its variants, coupled with gaze aversion from the prior speaker (*ibid.*). This 'multimodal assembly' serves as a preface to dispreferred responses across various sequence types. The use of 'I do not know' combined with gaze aversion is a routinized multimodal resource for prefacing dispreferred responses, extending beyond responses to polar questions and encompassing a wide range of sequence types, including responses to proposals, assessments, and informings (*cf.* Pekarek Doehler et al., 2021). These findings not only align with earlier studies linking gaze aversion with dispreference but also broaden the scope to demonstrate that this practice holds across a diverse set of languages, highlighting the universality and significance of this multimodal interactional strategy.

Pekarek Doehler et al.'s (2021) excerpts further demonstrate that recipients' gaze aversion from prior speakers typically occurs in the transition space or simultaneously with the response onset. Rarely does it take place during the preceding speaker's turn, and even then, it happens after the recognition point of the prior action and the conditionally relevant next action. Respondents also tend to return their gaze to the prior speaker toward the end of their responsive turn, marking a shift in engagement. They argue that these behaviors serve various functions: Gaze aversion can project a dispreferred response in a premonitory way, signaling the nature of the upcoming response. Conversely, maintaining gaze on a speaker can indicate an expectation of turn continuation, while mutual gaze between participants often signals alignment, even during moments of disagreement (*ibid.*).

However, Krug, 2025 found that participants use gaze aversion to signal self-involvement as a state of unavailability, managing interactional impasses resulting from disalingment. Krug observed that participants redirect their foveal attention to interactionally less relevant areas to avoid visually addressing other participants. The timing of gaze aversion in Krug's data aligns with the findings of the aforementioned literature, reinforcing its importance as a communicative practice.

In sum, these findings suggest that gaze aversion is a critical visual practice for displaying disalignment and projecting dispreferred responses. Nonetheless, research on gaze behavior during assessment sequences remains limited, as Krug, 2025 also notes for phases of disalignment, highlighting the need for further investigation. In this regard, Kendrick and Holler (2017: 18) emphasize that mutual gaze is more prevalent in face-to-face configurations, suggesting that spatial arrangements significantly influence gaze dynamics. This observation underscores the importance of considering environmental factors in understanding gaze behavior and its implications for social interaction.

Regarding gaze behavior in mobile settings, various studies have shown that speakers naturally spend less time gazing at each other compared to static settings (Auer and Laner, 2025; Laner, 2022; Auer and Zima, 2021; Stukenbrock and Dao, 2019). Thus, gaze is not the primary means to establish a sense of togetherness ("with," Goffman, 1971), but rather the synchronized moving together through space while walking alongside each other. Nevertheless, gaze remains an important interactional resource for response mobilization and monitoring, such as during question-answer sequences or sequences that include *laughables* (*cf.* Auer and Laner, 2025), meaning that participants do look at each other frequently, albeit for very brief periods, resulting in a low average duration of mutual gaze.

In summary, gaze behavior plays a multifaceted role in evaluative stance-taking sequences, influencing social dynamics and communication patterns. While much progress has been made in understanding its functions and implications, there remains ample room for further exploration, particularly in different settings (and different number of participants, as Rühlemann et al., 2019 propose). By delving deeper into the complexities of gaze behavior, researchers can enhance our understanding of human interaction and pave the way for more nuanced analyses of social dynamics. In this regard, the current study adds to this by exploring a setting that has not received much attention in stance-taking research, namely joint walks that occur mainly in side-by-side formations. This formation has also been described by Kendon (1990); however, gaze, and especially mutual gaze, naturally differ from vis-à-vis settings. For a more detailed discussion, see Auer and Laner (2025).

3 Data and method

The current contribution is based on ten videos from a corpus that was created within the research project Looking, Noticing, Talking: How Walkers Experience the Black Forest National Park[2] that consists of twelve video recordings, each approximately 90 min long. Participants walked with mobile eye-tracking glasses along a pre-determined route through the Black Forest National Park. Specific orientation points were given to the hiking pairs beforehand for navigation. The glasses allow for a precise analysis of the participants' gaze behavior, which is essential to reconstruct how joint attention is established and how objects in nature are noticed and evaluated. For sequence analysis, methods of interactional linguistics (Couper-Kuhlen and Selting, 2001, 2018) and multimodal conversation analysis (e.g., Mondada, 2014; Sidnell and Stivers, 2013) were applied.

To collect the present data, mobile eye-tracking glasses from Tobii[3] were used, which record both gaze behavior and conversations. An external camera was not used in order to create a more natural recording situation. After collection, the data from the hiking pairs were anonymized, synchronized, and split screens were created using Adobe Premiere Pro[4], in which the recordings of the two hikers are displayed side-by-side (see Figure 1). The red and green dots in the split screen show where the participants are currently looking. However, it must be noted that the front camera of the eye-tracking glasses does not cover the entire human field of vision (for a detailed discussion on the method, see Weiß, 2020: 38–43).

In a next step, all instances of assessments that occur in sequentially first positions were extracted and transcribed following GAT2 (Selting et al., 2009). The data excerpts used in this publication were further enhanced with still images and detailed non-verbal information, following the approach of Mondada (2017) and Merlino and Mondada (2019). Additional symbols above the verbal transcript illustrate the bodily orientation and gaze behavior of the interlocutors (see Figure 2), enabling a comprehensive multimodal analysis. For a detailed description of the conventions for the transcription of gaze *cf.* Laner (2022).

4 Analysis

4.1 Assessments in joint walks

In the current dataset, there are 98 instances of first assessments containing evaluative adjectives[5] in first positions that refer to entities in the surroundings, along with another 29 cases of general assessments of the surroundings (e.g., evaluating nature in general or commenting on the weather). Thus, assessments in the first position occur 127 times in the data during situated speech. Conversely, only 19 cases of assessments in first position can be found in the data during displaced speech. The study will focus on the cases during situated speech but will briefly discuss first assessments in displaced speech in Section 4.4 for comparison.

First assessments in the data occur in various formats (see below), including exclamatives (*cf.* in detail Auer et al., 2024). They may be prefaced for example by German perception imperatives 'guck' or 'schau' (both meaning 'look'), and/or interjections (*cf.* response cries Goffman, 1978), which can serve additional interactional functions (e.g., guiding the other's gaze or implicitly expressing affective stances before doing so explicitly). All the examples provided are extracted from the corpus and are produced as firsts. I will focus on the gaze patterns within these sequences.

(a) Oah kUck mal wie **SCHÖN**. (‚oah look PTCL how beautiful')
(b) !OH! SCHAU mal-= =ein **schÖner** FLIEgenpilz. (‚oh look PTCL a beautiful toadstool').
(c) **GEIL** die wurzeln wo die han; (‚sick the roots that they have')
(d) **TOLL** wie das Alles so (.) verMOOSt is auch; [ne?] (‚awesome how this is all full of moss, right?')
(e) dA **schöner** WEIHnachtsbaum, (‚there pretty christmas tree')
(f) sieht so **CRAzy** aus. (‚looks so crazy')
(g) die BANK is **sÜß** da; (‚the bench is cute there')
(h) der BAUM is **cOOL**. (‚the tree is cool')
(i) der schaut A **schön** aus; ge? ('that one also looks beautiful, right?')

The discussed literature in the background section suggests that there are two phases that become important for analyzing gaze behavior during assessment sequence: (1) the phase in which participants establish joint attention and one participant assesses the referent and (2) the phase in which the other participant responds to this first assessment. The stance triangle (*cf.* Du Bois, 2007: 163 and in this collection *cf.* de Vries et al., 2024) that is often discussed in literature can be applied to the current data as illustrated in Figure 3.

As shown, participants typically position themselves in a triangular formation toward the assessable. For the first part of the sequence, they adopt and maintain this position during the establishment of joint attention and the first assessment. During the second part (responding to the assessment), it is suggested that participants either turn toward each other and engage in mutual gaze (when showing agreement, *cf.* Haddington, 2006) or avert their gaze (when disagreeing with the initial assessment, *cf.* Robinson, 2020). This may be because, upon hearing an assessment, the recipient turns its gaze toward the person who made the initial assessment to show agreement. Consequently, Haddington (2006) argues that the participants are likely to look at each other during (agreeing) second assessments.

The gaze patterns in the data align with those described by Haddington (2006: 284) for the first part, including instances where walkers evaluate objects, they are both already looking at (see also Botsch et al., 2025). However, gaze behavior deviates from Haddington's observations in the second part, which becomes particularly evident in cases of disaffiliation as discussed in section 4.3.1. Note at this point that Haddington's findings are derived from static settings, where participants are seated around a table. This may be one factor influencing the differences observed. However, in the current context, the first part of the interaction typically occurs while

2 Cooperation of the Black Forest National Park, Department 3 (Dr. Kerstin Botsch, Dr. Susanne Berzborn), and the Chair of German Linguistics at the Albert Ludwig University of Freiburg (Prof. Dr. Dr. h.c. Peter Auer).

3 Tobii Pro Glasses 2: https://www.tobiipro.com/de/produkte/tobii-pro-glasses-2/.

4 Adobe Premiere Pro v14.0: https://www.adobe.com/products/premiere.html.

5 Instances in which stances towards natural objects were shared only implicitly (e.g., through laughter, etc.) are excluded from this investigation.

FIGURE 1
Split screen of the synchronized data.

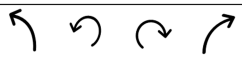
	Pentagons above the verbal transcript iconically represent the two walkers (with their initials) and their bodily orientation
	Arrows indicate the walkers' gaze directions
	Dashed arrows indicate that the person is gazing downwards
	Curved arrows indicate that the participants reorient their gaze and body in the direction of the arrow (e.g., turning towards the object of reference)
	Various icons represent the objects of reference. They are only represented in the transcript if at least one of the participants focuses on the object.
1 {verbal trans}{cript}	Curly brackets indicate the scope of the iconically illustrated gaze behaviour above the verbal transcript

FIGURE 2
Symbols used for gaze transcription.

participants are still walking, whereas the second part can take place during a stationary phase – i.e., when participants have stopped in front of the assessable, either during or after the noticing-assessment. This means that, in cases where participants halt to inspect a noticed object in nature and possibly agree or disagree with the initial assessment, there is little difference from stationary settings during this phase, as they can easily look at each other. Before delving into these gaze patterns, I will first provide an overview of how participants verbally react to initial assessments. Almost every second assessment is responded to in a very short and relatively neutral form (see Figure 4), such as with a simple "yes/yeah" or "mh_hm" (*cf.* acknowledgement tokens; Jefferson, 1984). Another 20% of assessments receive only non-verbal reactions, such as gaze and bodily positioning toward the assessed object[6] (*cf.* also Extract 5). Contrary

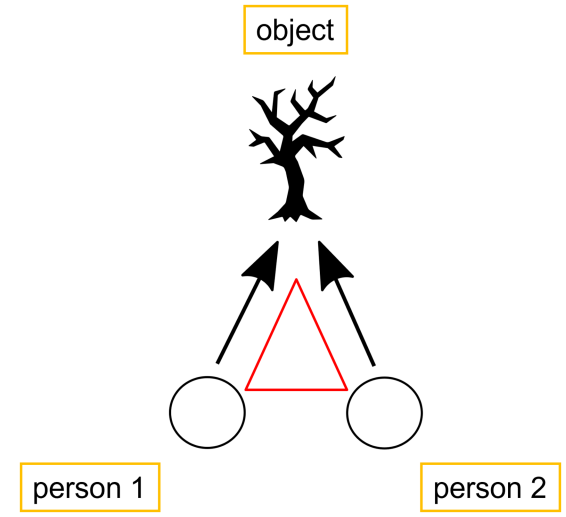


FIGURE 3
Participants gaze at an assessable in nature.

6 Note that very small additional non-verbal cues, such as facial gestures (e.g., smiling), might be visible to the speaker peripherally but will not be recorded by the eye-tracking glasses. However, these still fall into this category since there is no language involved.

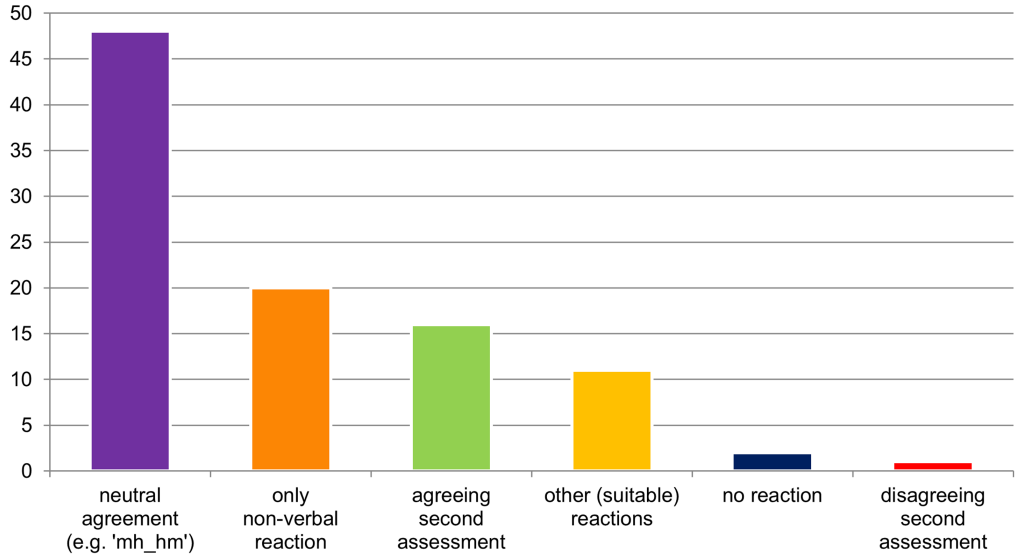


FIGURE 4
Distribution of different reactions to the assessments.

to expectations, only about one-sixth of the reactions are formulated as second assessments (approximately 16%, see above), with only one case contradicting the initial assessment.

Another way to react to the initial part of the assessing sequence is not to directly respond to the assessment itself, but to the object by asking more about it or expressing surprise upon seeing it (e.g., a rare plant). These cases are categorized as 'other (suitable) reactions.' For the sake of completeness, there are also two cases in which no reaction to the assessment follows. In these sequences, something else captures attention immediately after or almost simultaneously with the assessment, resulting in the assessment being neglected.

Altogether, it becomes clear that what is assessed as 'beautiful,' 'nice,' 'cool,' or 'sad' in nature is almost always agreed upon. This is likely due to cultural and social conventions that loosely define what is considered esthetic and what is not. This cultural imprinting, combined with the familiarity among participants, explains why disaffiliation – i.e. disagreement on the initial assessment – rarely occurs in the data.

4.2 Gaze patterns during affiliating assessments of objects in nature

As already discussed, Haddington (2006: 309) argues that participants first look at an assessable while one of them utters a first assessment and subsequently establish mutual gaze whenever the other participant adopts a convergent stance about the assessable. The act of jointly looking at the assessable typically begins before the actual assessment is uttered and is maintained for a significant period. In contrast, mutual gaze, which Haddington (2006) describes as occurring during the second part of the assessing sequence to display a convergent stance in response to the initial assessment, does not occur in the discussed data during a second assessment.

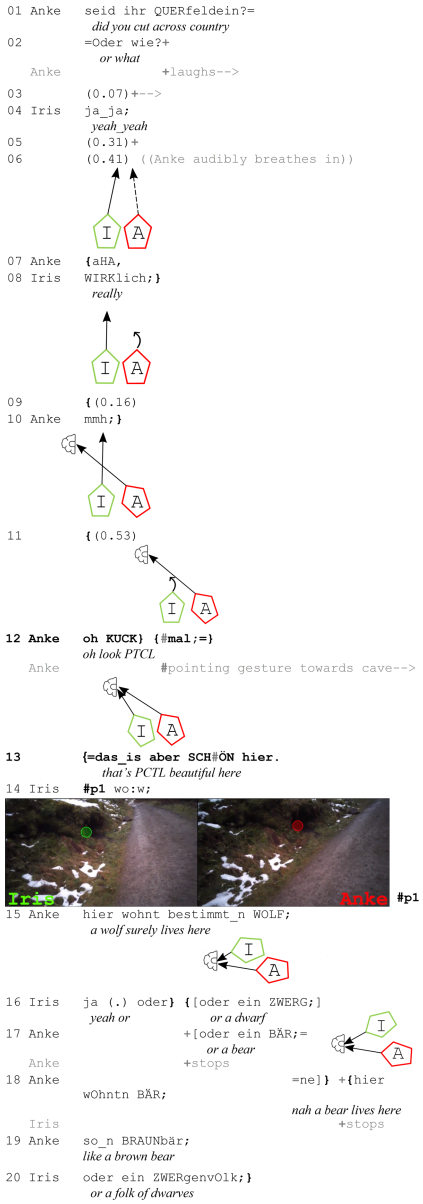
The activity of jointly looking at the assessable can be observed in each assessing sequence in the data. In the following section I will discuss two examples in which language is needed to introduce a new focus on an object that is assessed subsequently and two in which something is assessed that is already more or less in the focus of the participants, showing differences in the choice of utterances (the findings are consistent with Auer et al.'s (2024) findings for the use of how-exclamatives with or without a preceding perception imperative). In each case, the gaze behavior in the second part of the assessing sequence will be discussed in detail, too.

4.2.1 Establishing joint attention and assessing an object in nature subsequently

In the majority of cases (61%), one participant gazes at an object before calling attention to it and assessing it. This happens very quickly and precisely, rather than in a slow manner where joint attention is established separately from the assessment sequence. Eye-tracking data, which records gaze behavior precisely, shows that this process occurs sequentially and in an orderly fashion. In 21 cases, a perception imperative introduces the assessment. These imperatives can be uttered in its own intonation phrase or within the same phrase as the assessment, typically following one another without pauses in between. In all these cases, the gaze reaction usually begins immediately after the perception imperative is uttered (for a detailed discussion see Laner, 2022: 11). The following two excerpts illustrate this.

In the first example, the two walkers, Anke and Iris, are just concluding a conversation about a hike in Corsica. Following this sequence, attention is drawn to a small cave on the left side of the path. The small pentagons iconically represent the physical orientation of the participants, and the gaze behavior is depicted by arrows.

Extract (1): Little Cave (#Kleine Höhle; VP2122)



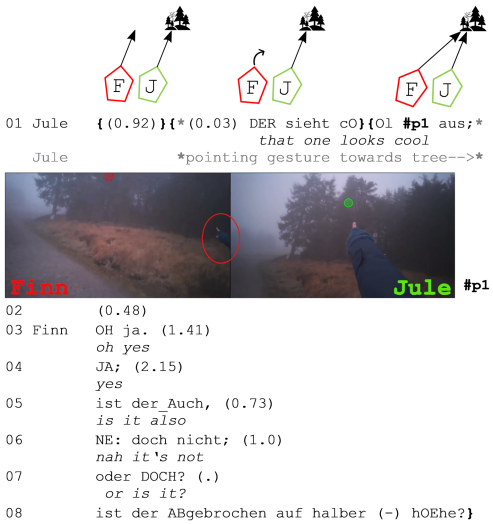
Toward the end of the preceding sequence, during which the two walkers talk about a previous hike, Anke turns to the left side of the path (line 09), with the gaze point from her eye tracking glasses showing that her gaze is directed at a small cave in line 11. A good half-second later, she then produces "oh KUCK mal;==das_is aber SCHÖN hier." (*oh LOOK PTCL== this is PTCL NICE here*; lines 12–13). Here, the perception imperative is preceded by the response cry 'oh' (*cf.* Goffman, 1978; Golato, 2012; Anna and Pfeiffer, 2021; Pfeiffer and Anna, 2021), which enhances the affective stance that is then made explicit through the assessment 'this is PTCL NICE here'. The addressee's bodily reaction occurs very quickly: Right after 'look' in line 12 and simultaneously with Anke's pointing gesture, Iris starts to refocus toward the space of reference. By the beginning of the assessment in line 13, her gaze is already directed at the targeted object. Thus, immediately after uttering the perception imperative (line 12), joint attention is established. Soon after, Iris lowers her hand and ends the pointing gesture. This occurs just before the end of her assessment in line 13.

As exemplified in the first part of this assessment sequence, the refocusing by the addressee (here line 12) typically begins before the place and/or object of reference is even mentioned or specified. Recipients, therefore, anticipate in this setting that 'look' is used in its literal sense (i.e., not as discourse marker, see Deppermann, 2021: 201, 203 and Günthner, 2017: 105) and they routinely align their gazes at the entity of reference. The effectiveness of this can be partially explained by the strong mutual monitoring (particularly through peripheral vision) that occurs while walking together (Deppermann and Schmitt, 2007: 121; Stukenbrock, 2015: 54), allowing participants to infer where the speaker is most likely referring to (for a more detailed discussion on the usage of perception imperatives in German see Laner, 2022).

For the second part of this sequence, Iris verbally reacts to the assessment with "wow" (line 14), indirectly acknowledging that she has recognized the object of reference and affiliating herself with Anke's positive assessment (on alignment and affiliation in more detail see, among others, Stivers, 2008; Steensig, 2019). During this phase, they both continue to gaze at the assessable and even afterwards. There is not a single gaze toward each other, even though they have stopped to inspect the little cave. This suggests that, even though they have entered a stationary phase – comparable to stationary settings as observed in Haddington (2006), Kendrick and Holler (2017), and Pekarek Doehler et al. (2021) – they do not establish mutual gaze during the affiliating second assessment. Only Haddington included assessments of objects in the current surroundings, though, which results into building up a stance triangle (Du Bois, 2007: 103) with a visible object as shown in Figure 3.

In the second example, the two participants are in an open state of talk, before one of the walkers starts to point at a tree on the right side of their hiking trail and assessing it.

Extract (2): Tree (#Baum; VP2728)



In this case, there is no perception imperative preceding the assessment, but a pointing gesture begins before the utterance, which is rare in the data (they usually start later). As seen in the still from Finn's camera (#p1, left side), the pointing gesture is clearly visible in his right periphery. This likely explains his early gaze reaction (turning toward the object) at the start of Jule's assessment (line 1). When Jule utters her assessing adjective "cool" (line 1), both are already gazing at the assessable. After a short pause (line 2), Finn reacts with "oh ja" (*oh yes*, line 3), agreeing with her assessment of the tree. They then discuss whether the tree is broken at medium height, but this part of the discussion is excluded for reasons of space.

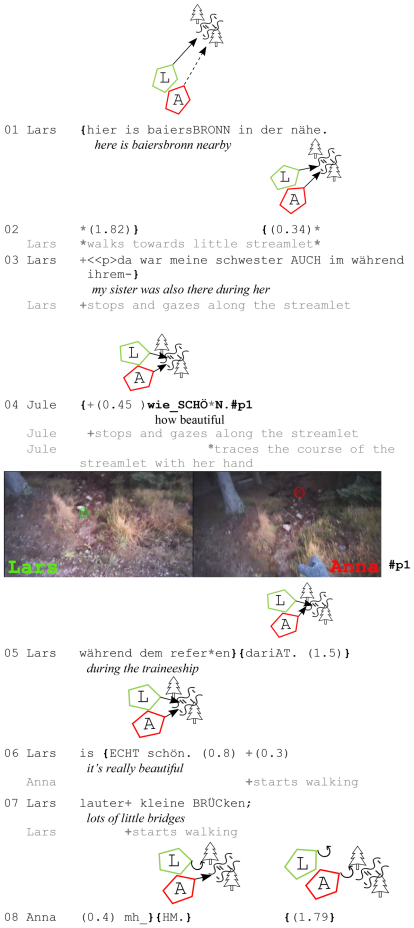
As in extract (1), joint attention is established right before the assessment is made explicit ("schön," *beautiful* in extract 1 and "cool" in

extract 2). Thus, the same gaze pattern can be observed during the first part of the sequence. The reactions in these two sequences differ in that the first is responded to with a second assessment ("wow"), whereas the second is responded to with "oh ja" (*oh yes*), which also conveys (affective) agreement (see Golato, 2012 for a discussion of "oh" as an affective change-of-state token). Despite the slight difference in reactions, no difference in gaze behavior is observed in the second part of the sequence: Both participants only gaze at the assessable and never at each other. In the next part, I will discuss two examples in which the gaze behavior in the first part of the sequence differs (a joint focus on the assessable is already established before the sequence with the assessment even begins).

4.2.2 Assessing an object in nature that is already in the focus of both walkers

As Botsch et al. (2025) argue, joint attention (or at least a joint focus) can be inferred by two walkers from their bodily behavior. This becomes especially clear when both walkers stop (more or less together) to gaze at an object in nature. Here again, a subsequently ordered gaze pattern can be observed, but for the first part, no verbal exchange is needed: Both participants orient toward the later assessed object and gaze at it (and stop to do so, if it is not an object far ahead). In the example below, the two participants are talking about a town nearby when they both stop and gaze at the course of a small stream.

Extract (3): Streamlet (#Bachlauf_1; VP0506)



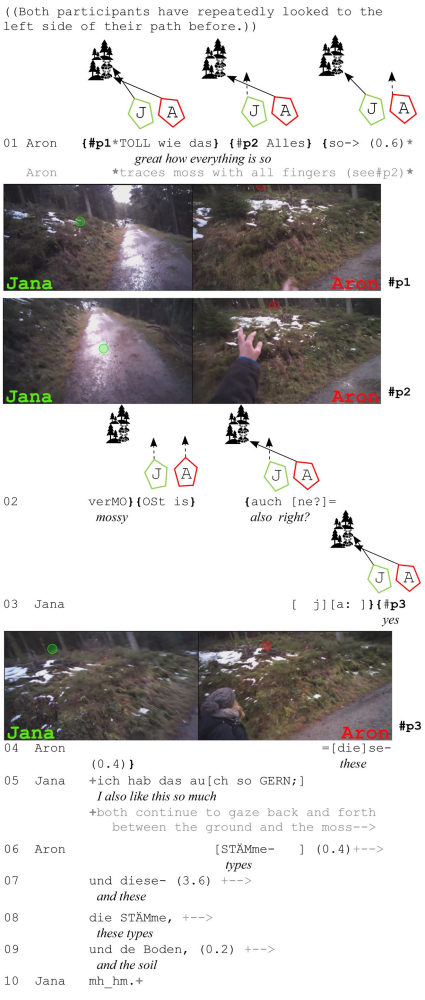
While Lars makes his reference to Baiersbronn in line 1, he starts to gaze at a small stream and its course. Shortly after (line 2), Anna also gazes at it. Both stop subsequently (Lars in line 3 and Anna in line 4) before Anna assesses it with "wie SCHÖN" (*how beautiful*, line 4). As both walkers have stopped and are gazing at the streamlet, it can be assumed that they are both aware that they are looking at the same object [*cf.* Botsch et al. (2025)]. "Wie schön" (*how beautiful*) here can be described as a minimal form of a how-exclamative (see Pfeiffer, 2017: 43). The reduced form of "wie" (*how*) and the evaluative adjective "schön" (*beautiful*) strongly supports the assumption that both participants have already accomplished a common visual focus that is presupposed by both of them [as argued by Auer et al. (2024): 269f, extract 6].

The observable gaze pattern during the first part of the assessment sequence is in this case similar to the one seen in the first two extracts, but without verbal exchanges being necessary. Although joint attention is already established, Anna initiates an iconic pointing gesture toward the end of her assessment, tracing the course of the streamlet. This gesture highlights that she finds the course of the streamlet beautiful, not just the streamlet itself.

Lars finishes his utterance about his sister in line 5 before he reacts to Anna's assessment with an upgrade: "is ECHT schön" (*it's really beautiful*, line 6). Here again, the two walkers stay in their triangular position facing the assessable without once glancing at each other.

In the next example, both walkers are also gazing at the assessable before the assessing sequence starts. In contrast to the example before, the participants do not stop in this extract, but continue to walk while they talk about the moss that can be seen everywhere on the left side of the walking trail.

Extract (4): Moss (#Moosstämme; VP1920)



In this case, it is important to emphasize the bodily positions of the walkers. Jana walks on the right side and because of Aron's orientation toward the left side, she is now even more perceivable in his periphery. Thus, he can perceive that she is also looking at the left side and seems to infer she is also gazing at the moss on this side of the path, since he neither produces any perception imperatives nor uses any initial (local) deictic terms. He starts with an assessing adjective[7], but still uses an iconic gesture (see #p1) – as in the example before – to illustrate his point. Jana agrees with "ja" (*yes*) and continues to explain that she likes that a lot too.

In the extracts discussed so far, I have explored the intricate relationship between gaze behavior, joint attention, and assessments in mobile settings, revealing a distinct pattern of gaze use during the noticing-assessment process. The first two examples clearly demonstrate a sequential gaze pattern, where both participants direct their gaze toward the object to be assessed before the assessment is verbally articulated. This pattern is consistent across various extracts, with the only variation being whether the initial gaze occurs before the verbal utterance [extracts 3 and 4, *cf.* Botsch et al. (2025)] or during the noticing-assessment (extracts 1 and 2). Despite this variation in timing, the second part of the assessment sequence remains consistent: participants continue to gaze at the object while agreeing (affiliating) with the initial assessment. This continuity suggests that gaze serves as a fundamental tool for establishing joint attention during the whole noticing-assessment sequence, regardless of when it is initiated and how participants react to it.

This observation is crucial because it challenges assumptions about the role of gaze in signaling affiliation. In many other contexts, mutual gaze is an important marker for agreement (or affiliation), as seen in the work of Haddington (2006), Kidwell (2006), Kendrick and Holler (2017), and Pekarek Doehler et al. (2021) among others. These studies highlight how mutual gaze can serve to visibly signal a convergent stance and mutual agreement. However, in the current setting of mobile side-by-side interactions, we find a significant departure from this pattern. Contrary to what might be expected based on previous research, walkers in this study do not engage in mutual gaze to signal their agreement with the other participant's assessment. Rather, gaze is used primarily to establish joint attention, not to indicate affiliation.

This raises the question: do gazes toward each other simply not occur during mobile side-by-side interactions, or do they serve a different function altogether? As will be shown in the following section, participants do indeed gaze at each other in these contexts, but for reasons that differ significantly from signaling affiliation and agreement. This distinction is important, as it suggests that gaze serves a more fundamental role in ensuring attention is directed toward the object of assessment, rather than fostering mutual agreement between participants.

Drawing on Haddington's (2006) work, we can understand this pattern more clearly. Haddington argues that participants first look at an assessable while one of them makes an initial assessment, and mutual gaze typically emerges during the second part of the assessing sequence, when the other participant adopts a convergent stance toward the object. In this sense, mutual gaze functions as a sign of affiliation, showing that both participants are in agreement with the assessment. However, in the data discussed here, mutual gaze does not occur during the second agreeing assessment. Despite having entered a stationary phase, which could be compared to the stationary settings described by Haddington (2006), Kendrick and Holler (2017), and Pekarek Doehler et al. (2021), participants do not establish mutual gaze during the second assessment.

This divergence suggests that gaze behavior during mobile interactions functions differently than in stationary settings. While mutual gaze may be used to display affiliation in stationary settings, in mobile contexts, it serves more to establish joint attention on the object being assessed, rather than signaling agreement between the participants themselves. This shift in function highlights the adaptive nature of gaze in different interactional settings, where one of its roles in one context – such as signaling affiliation – may change in another context, where its role is more focused on directing attention to the object of assessment.

Moreover, only Haddington's study included assessments of objects in the immediate surroundings, which contributes to the formation of a stance triangle (Du Bois, 2007: 103) involving the two participants and the object that is assessed. This triangle (Figure 3), is critical to understanding the unique nature of gaze behavior in these contexts. The presence of a visible object in the current surroundings plays a central role in the interaction, influencing how participants engage with it and each other. While mutual gaze may emerge in stationary settings as a way of displaying affiliation, in mobile settings, it seems that gaze is less concerned with signaling agreement and more focused on ensuring that both participants share attention toward the object being assessed – even during stationary phases, whenever participants have stopped in front of the assessable as in extracts 1 and 3.

In conclusion, these findings suggest that gaze behavior in mobile interactions functions primarily as a tool for establishing joint attention rather than signaling affiliation. This insight challenges previous assumptions about the role of gaze in social interaction and calls for a more nuanced understanding of how gaze operates across different interactional settings. By focusing on the ways in which gaze facilitates shared attention to the object of assessment, we gain a deeper appreciation of its role in mobile social contexts, where the dynamics of interaction differ from those in stationary settings (*cf.* also Auer and Laner, 2025). Further research is needed to explore the subtleties of gaze behavior in mobile interactions and to better understand how gaze functions in other contexts than the typically investigated sitting arrangements.

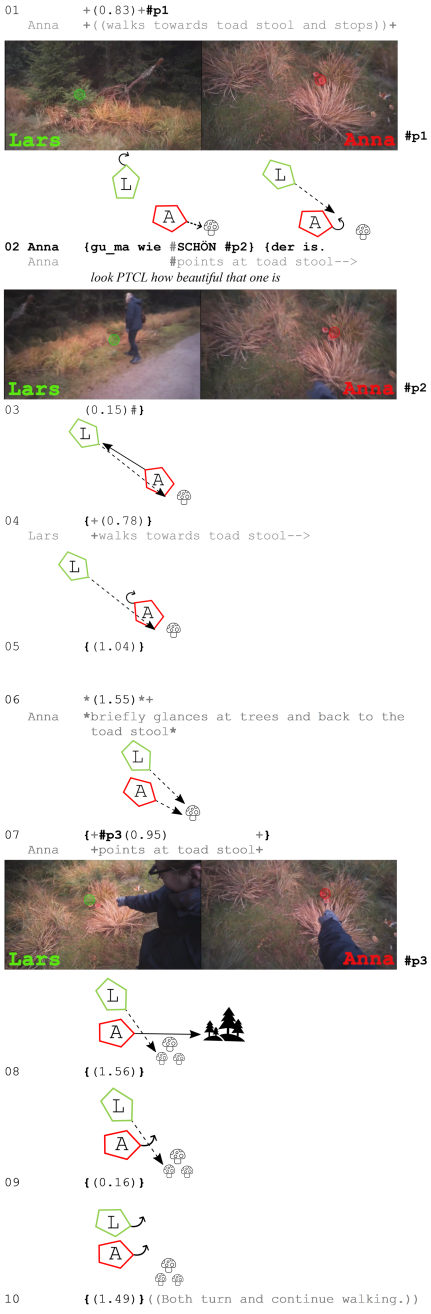
4.3 Gazing toward each other during assessment sequences

As shown in the extracts before, assessment sequences that proceed in a regular manner, i.e., in a side-by-side configuration and with preferred reactions to the first assessment, the two walkers never gaze at each other. However, there are seven cases in the data in which participants gaze at each other. In this section I will first show an example in which the speaker of the assessment gazes at its recipient, before I will show an example in which the recipient of the assessment gazes at the speaker. I will then conclude in the last section (4.3.1) with an extract in which both walkers gaze at each other, i.e., establish mutual gaze.

In the following example, the two walkers are in an open state of talk when one of them perceives a toad stool on the right side of the path which will be assessed in the following.

7 Note that this structure of utterance (assessing adjective in first position) only occurs in the data when entities in nature are assessed that are not one singular object (e.g., snow, forest, moss).

Extract (5): Toad stool (#Fliegenpilz5_1; VP0506)

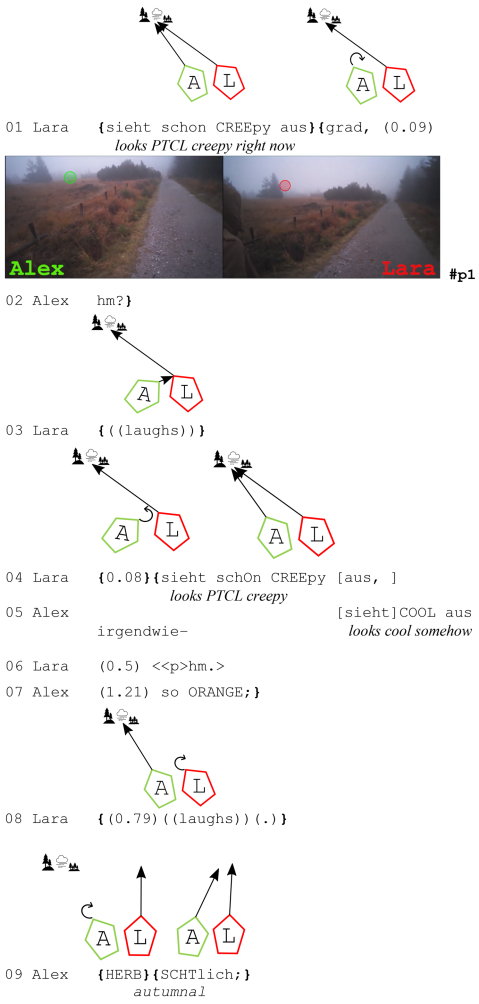


Before Anna assesses the toadstool as beautiful, she first walks toward it and stops right in front of it. This behavior has two consequences: First, she has moved away from her walking partner (she has fallen behind) and therefore cannot perceive him in her periphery anymore (she has even turned her head completely away from Lars, as can be seen in #p2). Second, Lars seems to recognize that she has stopped and fallen behind because he already starts to turn around at the beginning of line 2, right before she starts her utterance. He follows her gaze and looks at the assessable shortly afterward (also line 2). However, toward the end of Anna's assessment, she turns completely around to look at Lars. She gazes at Lars during line 4, when he starts to walk (back) toward her and the toadstool. She then turns back to the assessable again (line 5) and waits for him while briefly gazing at the trees and back to the toad stool (line 6). In line 7, she presumably perceives that he is now next to her in her periphery, because she then leans forward and produces another pointing gesture (#p3). They both continue to gaze at the toadstool before subsequently turning and starting to move again together.

In this example, the speaker of the assessment gazes toward her recipient. This happens already during the uttering of the assessment and only in cases in which the participants have separated from each other or the referent is especially hard to 'detect' and a lot of direct monitoring and explaining is needed to show the other the assessable. Cases like these are rare (only 4 cases in the data set), but they show that participants do look at each other during assessments, but for other reasons than affiliation – in this case the monitoring function of gaze is crucial.

The next extract shows one of the two cases in the data in which the recipient of the assessment gazes toward the speaker (so the other way around to the case before). As in the extract before, the two participants are in an open state of talk before the assessment sequence.

Extract (6): Foggy Nature (#Natur_Nebel; VP0304)



Here, both participants gaze at the landscape covered in fog before Lara utters, "sieht schon CREEpy aus grad" (*looks pretty creepy right now*, line 1). Alex does not understand her and initiates repair with "hm?" in line 2. She then laughs and repeats her assessment of the foggy landscape (lines 3 and 4). During her repeated assessment, Alex assesses the nature as 'somehow cool' (line 5) because the leaves are "orange" (line 7), and it looks "herbstlich" (*autumnal*, line 9). Although at first glance it might seem that the assessments diverge, both assess slightly different aspects of what they are gazing at in this moment, but

still show agreement (e.g., Lara utters "hm" in line 6). Additionally, they have repeatedly talked about this during their hike that the foggy landscape looks a bit 'creepy' but also 'cool' at the same time.

Here, a gaze toward the speaker of the first assessment can be observed. This happens during the repair initiation and thus apparently is not connected to the assessment itself. This gaze behavior during repair initiations is also found in other sequences in the data that do not contain assessments.

As these two extracts (5 and 6) have shown, whenever one of the participants gazes toward the other, it happens for monitoring reasons (speaker toward recipient, extract 5) or while initiating repair (recipient toward speaker, extract 6). These gaze orientations toward the other walker are not connected to the assessment sequence but have different interactional reasons. However, there is one case in which participants turn toward each other and even establish mutual gaze because of what happens during the assessment sequence. This will be discussed in the following subsection, 4.3.1.

In the previous chapter, I observed that gaze behavior during assessment sequences primarily serves to establish joint attention to the assessable object rather than signaling affiliation between participants. Mutual gaze, a typical marker of agreement or convergence in other contexts (Haddington, 2006; Kendrick and Holler, 2017; Pekarek Doehler et al., 2021), was notably absent during the assessment process, particularly during the second part of the assessing sequence, when participants are expected to agree with the initial assessment. Instead, the gaze was focused on the object, reflecting the participants' shared attention to the object of evaluation, rather than engaging in mutual gaze as a sign of affiliative alignment.

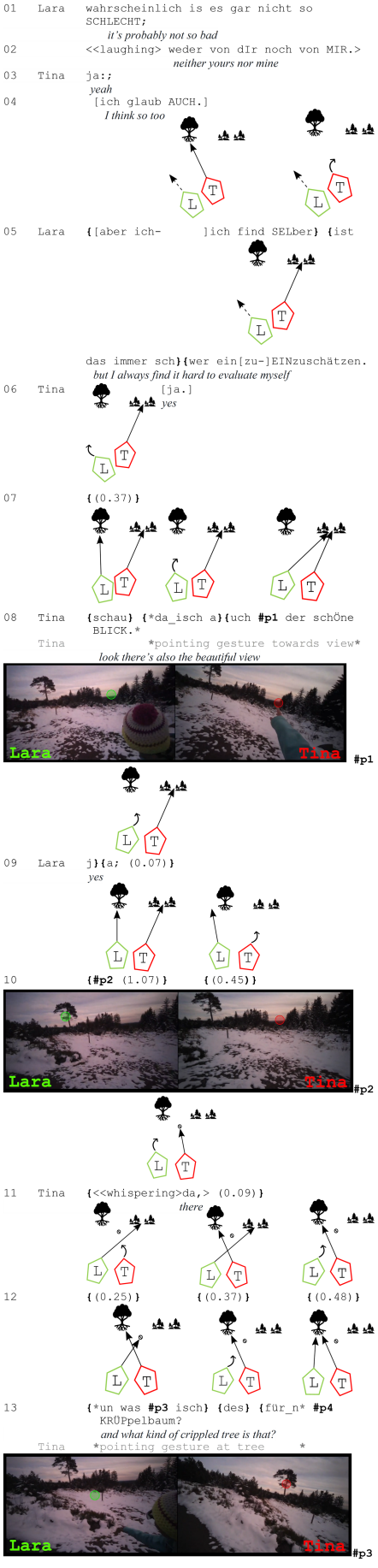
As the two extracts (5 and 6) in this section have shown, whenever one of the participants gazes toward the other, it serves monitoring purposes (speaker toward recipient, extract 5) or occurs during repair initiation (recipient toward speaker, extract 6). These gaze orientations are not connected to the assessment itself but serve different interactional functions. However, there is one case in which mutual gaze occurs as a result of the assessment process. This finding raises important questions about the conditions under which gaze behavior shifts from being a tool for monitoring and joint attention to other functions. It also offers a nuanced contribution to the literature on gaze dynamics in interaction (Haddington, 2006; Kendrick and Holler, 2017; Pekarek Doehler et al., 2021). In this instance, mutual gaze emerges not as a sign of affiliation, but rather, it is used in the context of strong disagreement with the initial assessment. This provides a rare example of how gaze can shift in response to the content of the assessment. This shift contrasts with earlier findings, where gaze toward each other was not part of disaffiliating second assessments, underscoring the complex, context-dependent nature of gaze behavior in mobile interactions.

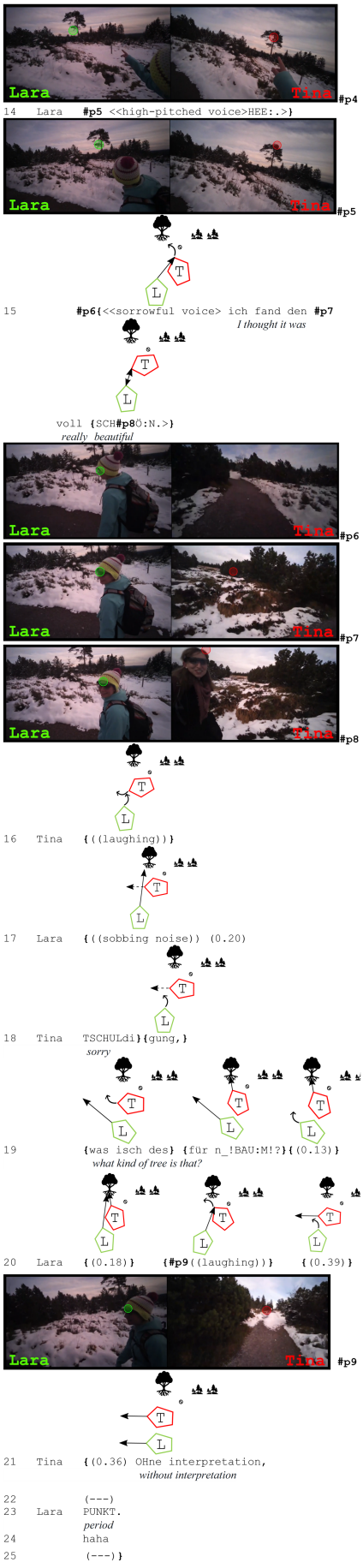
4.3.1 Mutual gaze during assessment sequences

As mentioned before, it is very rare for walkers to disagree on how nature is assessed, and it is even more unlikely for them to strongly disagree (see overview in Section 4.1, Figure 4). However, in cases where one walker strongly disagrees with the other's evaluation of an object, a different gaze pattern emerges during the second part of the sequence. The following excerpt shows this in detail.

In the beginning of this sequence (lines 1–7), they talk about something not connected to nature before one assesses the view (line 8) and subsequently a tree (line 13).

Extract (7): 'Crippled tree' (#Krüppelbaum; VP1516)





The excerpt above contains two assessments – one of the view and one of a tree. In the first case, Tina assesses the view as beautiful (line 08) and Lara agrees (line 09). Here, they both establish joint attention but do not look at each other (as discussed in the first part of section 4.2). During the second assessment, when Tina expresses her stance toward a tree by evaluating it as 'crippled' (line 13), they also establish joint attention: Tina points at the tree and formulates her assessment as a question "un was isch des für_n KRÜPpelbaum?" (*and what kind of crippled tree is that*, line 13). Up until this point, gaze behavior corresponds to the first examples (1 and 2).

Lara reacts with a high pitched "hee:" (line 14) that sounds almost reproachful and starts gazing at Tina (see #p6). This 'he' projects her upcoming disagreeing assessment of the object (similar to prefaces such as 'I do not know' or 'well', *cf.* among others Robinson, 2020, Pekarek Doehler et al., 2021, and Heritage, 2015). She then utters in a sorrowful voice "ich fand den voll SCHÖN." (*I thought it was really beautiful*, line 15). During this second assessment, Tina also turns, and they establish mutual gaze at the end of the second assessment (#p8). As they turn away from each other again, Tina laughs (line 16) and Lara mockingly makes sobbing noises (line 17). Then, Tina apologizes (line 18) and reformulates her question, leaving out her negative assessment: "was isch des für_n BAUM?" (*what kind of tree is that?*, line 19) to which Lara reacts with laughter and a brief glance at Tina (line 20). Tina continues with "OHne interpretation" (*without interpretation*, line 21), and Lara concludes with "PUNKT" (*period*, line 23) and laughs again (line 24). Note that all of this takes place while they are walking.

Strongly disagreeing on how an object is assessed can be potentially face-threatening and requires careful negotiation to manage the social dynamics of the interaction. In such cases, participants employ various conversational strategies to mitigate the impact of disagreement as can be seen in this extract: (1) Laughing here serves as a softening mechanism, reducing the tension associated with disagreement and signaling that the disagreement is not meant to be confrontational (line 16, 20, 24). (2) Apologizing as in line 18 can serve to acknowledge the other's perspective and mitigates the impact of the disagreement. (3) Reformulating (line 19) can further help soften the disagreement, allowing participants to maintain a cooperative atmosphere. On top of all these strategies, mutual gaze (4) seems to play a crucial role in these negotiations, as it facilitates a more immediate and personal connection, enabling participants to gauge each other's reactions and adjust their responses accordingly. By combining these strategies, participants navigate the complexities of disagreement, maintaining respect and cooperation despite their differing views.

In summary, the last three examples demonstrate that when two people assess things in nature while walking together, they do not simply avoid gazing at each other. Instead, they gaze at each other for specific reasons that are not necessarily connected to the assessment itself. Speakers tend to gaze at recipients when the two walkers are physically distanced from each other, requiring the speaker to monitor the recipient's position and reactions. Recipients, on the other hand, gaze at speakers during the shown assessment sequences primarily when initiating repair during conversation. Mutual gaze, distinct from these other

forms of gaze behavior, occurs due to divergent assessments. When participants strongly disagree, mutual gaze is established as a way to negotiate and resolve the disagreement, likely because such disagreements can pose a face-threatening situation.

4.4 Gaze during first assessments in sequences of displaced speech

Although sequences of displaced speech are not the focus of this paper, I would like to briefly address them in this section to provide a broader perspective. Assessments in first position during displaced speech are rare, as they typically occur in second position, e.g., when reacting to a story or something similar. In the current dataset, there are 18 instances of utterances containing assessments in a sequentially first position. Four are used to frame a following story, projecting what kind of story it is, e.g., 'funny', 'sad', etc. [*cf.* examples (d) and (e)], and two are used to resume a topic previously set aside when participants shifted from displaced to situated speech (e.g., when they noticed something in nature while discussing a topic not related to the current surroundings, *cf.* example f). The other twelve instances are first assessments that introduce a new topic and assess a referent in the same utterance (*cf.* examples a-b) and can either be assessables known to both participants (as in b and c) or not (as in a).

(a) wir ham da letzte STEINpilze gefunden die war_n richtig geil;
('*we found some porcini mushrooms there last time; they were really awesome*;')
(b) <<lachend>die HANna vorhin war au lustig ge,>
('*<<laughing>Hanna was also funny earlier, right,>*')
(c) EOH; (-) KRASS,= =nAEchsten SONNtag ist schon erster advent;
('*EOH; (-) SICK,= =next Sunday is already the first Advent*;')
(d) das war gestern LUStig,=
('*that was funny yesterday*,=)
(e) äh:m (-) auf der HOCHzeit was_n bisschen n fAIL war-
('*uh:m (-) what was a bit of a fail at the wedding*-')
(f) ja also richtig ver[RÜCKT.]
('*yeah, really crazy*.')

In all these sequences, no differences in gaze behavior were found compared to the cases discussed earlier during situated speech, even though in these instances, no stance triangle (Du Bois, 2007) with a visible object in the surroundings needed to be established – i.e., no joint attention to an entity in nature had to be established. In all these sequences, no differences in gaze behavior were found compared to the cases discussed earlier during situated speech. To illustrate this, I will present two short examples. In the first extract, both walkers are silently walking next to each other before the sequence starts.

Extract (8): Porcino Mushrooms (#Steinpilze; VP0102)

01 Ella wir ham da letzte STEINpilze **#p1** gefunden=
we found porcino mushrooms last time
02 =die war_n ECHT geil; (---)
they were really awesome

#p1

03 Nina in ESCHbach?=
in Eschbach?
04 Ella =ja; (.)
yes
05 bei_der STERNwaldwiese; (.)
at the sternwald meadow
06 Nina **#p2** ach COOL; (-)

#p2

07 Ella warn mEga VIEle;
there were a ton

The shown sequence starts with Ella's utterance in line 1 and 2 that she has found porcino mushrooms which were "ECHT geil" (*really awesome*, line 2). Nina asks subsequently if she found them in Eschbach, which Ella affirms (line 4) and specifies with "bei_der STERNwaldwiese" (*at the sternwald meadow*, line 5). Only then Nina reacts with a second assessment (line 6), which does not assess the mushrooms itself, but the fact that Ella found them. The sequence ends with Ella saying that there were "a ton" of them there (line 7). Here, the assessable that Ella refers to is not accessible to Nina and can therefore only be assessed by Ella. Meaning that only she has the right to assess it (as only she has knowledge about it), which goes hand in hand with only her being able to assess it. Still, Nina produces a second assessment, but refers to the finding of them in Eschbach.

During this short assessment sequence, both participants continue to walk and gaze down at the ground (as in #p1). Their gazes do not remain fixed on one point but wander around the path they are walking on (compare #p1 and #p2). There are no gazes toward each other, and there is no triangular positioning since the assessable is not visible in the current surroundings. Similarly to the cases during situated speech, there are generally no gazes toward each other. However, some cases do involve gazes toward each other, but again for different interactional reasons, as I will shortly show in the last extract.

In the last extract, both participants are again in an open state of talk, before the sequence starts. Here, one participant assesses something that both have access to, i.e., something they experienced together.

Extract (9): Rope up (#Anseilen; VP0910)

01 Gisa <<laughing> die **#p1** Tessa vorhin war au lustig ge,
Tessa was funny too earlier, wasn't she?

#p1

02 (0.5)
03 *die hat des geGLAU*+BT,>
she believed that
Gisa *turns to Hans* +gazes at Hans' face-->

#p2

04 Gisa [((lau+ghs))]
05 Hans [((laughs))]
06 (1.7)
07 Gisa ((clears throat))
08 (1.4)
09 Hans ich glaub UNS glaubt se so manches.
I think she believes quite a few things from us
10 (1.0)
11 in DER richtung. ((la[ughs))]
in that regard
12 Gisa [((laughs))]
13 hast dein SEIL da<<laughing>bei,>
do you have your rope with you
[((laughs))]
14 Hans [((laughs))]
15 n_paar karaBIner; (-) ha;
a few carabiners
16 Gisa <<laughing> n_kurzes stück müsse ma uns ANseiln.>
a short section we need to rope up
17 ((blows her nose))
18 auf der WANderung;
on the hike
19 (0.4)
20 <<highpitched voice> WAS?>
what
21 Hans ((laughs))

In this sequence, Gisa recalls a funny joint experience, namely that they pulled Tessa's (a friend of their daughter) leg, by telling her that they would have to rope up during their hike today. Her assessment of Tessa is accompanied by laughter and ends with a question tag (line 1). Similar to findings for "ne" in German, here "ge" (both can roughly be translated with 'right'/'wasn't it' etc.) is used to indicate that Gisa does not claim exclusive rights to the assessment, showing that they have equal or shared epistemic rights to assess the referent (*cf.* König, 2017: 245). Then there is a short pause in line 2, before Gisa continues with "die hat des geglaubt" (*she believed that*, line 3) and turns toward Hans. Since the joint experience is also a laughable, the gaze behavior falls in line with Auer and Laner's (2025) findings that speakers regularly gaze at the recipients of their laughable during joint walks. She stops gazing at Hans right after he starts to laugh too.

During the assessment, both participants continue to gaze toward the path they are walking on, as in the previous extract (extract 8). There is only a speaker gaze later that seems to elicit a response, as the gaze stops immediately after the recipient starts to react with laughter.

In this section, the focus was on instances where the assessable is not a visible object in nature, but rather referents that are not present, such as people or objects that are not immediately visible. These cases, although similar to the findings from Kendrick and Holler (2017), Pekarek Doehler et al. (2021), and Siromaa and Rauniomaa (2021), differ primarily in terms of mobility. While the general verbal exchanges are (better) comparable to these studies, as there are no visible referents, the gaze behavior still differs and may be due to the mobility of the participants during the interaction.

However, as demonstrated by Auer and Laner (2025), gazing toward each other does occur during displaced speech, but for different interactional reasons. In contrast to findings in previous studies where mutual gaze often indicates agreement or shared stances (Haddington, 2006; Kendrick and Holler, 2017), in the case of displaced speech, gaze is not used to display affiliation. Instead, it serves different interactional functions, primarily eliciting responses from the other participant. During situated speech, gaze toward each other was similarly not tied to the assessment itself or used to indicate affiliation. Rather, gaze in these contexts functions more as a tool for managing the flow of interaction, establishing joint attention and ensuring that the recipient is engaged and responsive.

Thus, while gaze plays an important role in both types of speech, it does not serve the affiliative function often described in stationary settings. Instead, it is mobilized to manage the interaction and to elicit responses, underlining the complex, context-dependent nature of gaze behavior in dynamic, mobile settings. This distinction between gaze as a tool for joint attention and gaze as a tool for signaling affiliation highlights the different interactional demands at play when assessing not only visible but also displaced referents.

5 Conclusion

The present study examined interactions among people walking together to understand the gaze patterns during assessments of entities in the current perceptual space. The goal was to use a conversation-analytic and qualitative approach to identify and explore gaze patterns in each assessment sequence found in the dataset and providing a quantitative overview of these findings. Consistent with existing research focused on interactions while walking, particular importance was placed on bodily-visual practices in the examined sequences, with a special focus on gaze behavior. This study addresses a niche area of research, adding insights into the dynamics of mobile interactions. The conversation-analytic and qualitative approach allows for a detailed understanding of how walkers navigate the dual tasks of walking (and stopping) together while responding to and evaluating environmental stimuli.

Researching this particular setting has revealed that, contrary to existing literature, mutual gaze is not used to mark a "convergent stance" (Haddington, 2006: 299). Instead, the opposite was observed: walkers gaze at each other when they strongly disagree on an assessment of an object in nature. This highlights the social aspect of gaze, as strong disagreement is a face-threatening act (with laughter also playing a mitigating role, as seen in Extract 7). Since there was only one strong disagreement found in the data (although 127 cases of assessments occurred and were investigated), further research is needed to strengthen this finding. Still, the fact that in all the cases where participants agree, the described triangular position was not only held during the first part of the sequence but also during the second part (the response), strongly suggests that the established mutual gaze during the one case of

disaffiliation is significant. Furthermore, the establishment of mutual gaze happens while the participants continue to walk, whereas gazing toward each other would presumably be easier if they had stopped, as they did in Extracts (1), (3) and (5).

Thus, in the context of walking together through nature, gaze does not serve as a tool to signal affiliation during assessment sequences. Instead, the primary role of gaze in this setting is to establish and maintain joint attention on the surrounding environment and to ensure safe navigation along the path. The need to focus on the terrain and potential obstacles necessitates that participants direct their gaze toward the path, reducing opportunities for mutual gaze. This practical consideration overrides some of the communicative function of gaze seen in other settings, where for example mutual gaze can signal affiliation or alignment. Consequently, during walking interactions, joint attention to the environment becomes more critical, with participants using their gaze to visually engage with their surroundings rather than each other. This shift underscores that in such mobile side-by-side interactions, the coordination of physical movement and shared environmental focus takes precedence over (some) functions of gaze typically observed in more stationary contexts (and vis-à-vis). It is important to stress, though, that not all functions of gaze known from other settings are different, as seen in extract 6, where one participant initiated repair. In situated speech sequences establishing joint attention is presumably the most important function of gaze (also to monitor each other as in extract 5). During displaced speech, joint attention is not a factor, and therefore other functions of gaze can be found, as Auer and Laner (2025) show.

In addition to these primary findings, the study revealed several notable side findings. Almost half of the assessments were reacted to with short and rather neutral verbal reactions such as "mh_hm" or a simple "yes." Even more remarkable, 20% of the cases were reacted to only non-verbally, as shown in Extract (5), where a clear reaction is demonstrated by gazing at and walking back to the assessable – but there was still no verbal reaction. Contrary to expectations, second assessments occurred in only 16% of cases. These findings, apart from gaze behavior, are peculiar and presumably connected to the specific setting of the study. Naturally, in a setting where assessments occur so frequently (mean = 14.1 per hike), not every assessment can lead to longer sequences, but is rather quickly acknowledged before continuing to walk. Either way, further research in comparable settings (such as walks in different contexts) is needed to show whether these findings only apply to walks through nature, or whether mobility is the important factor.

In conclusion, mutual gaze does not play a significant role in assessments during mobile side-by-side interactions. Only one instance found in the data is (somehow) related to assessments: mutual gaze during disagreeing second assessments. This behavior indicates that mutual gaze is employed as a strategy to navigate potentially face-threatening moments and negotiate diverging stances. However, further research is necessary to determine if engaging in mutual gaze during mobile interactions is consistently linked to disagreement and face-threatening situations. Understanding these nuances can provide deeper insights into the intricate dynamics of gaze behavior in mobile contexts and its broader implications for social interactions.

Data availability statement

The datasets presented in this article are not readily available because written consent by participants was given to the storage and usage of the data at the University of Freiburg and publications, but not for providing the raw data (the videos) for others. Requests to access the datasets should be directed to barbara.laner@germanistik.uni-freiburg.de or peter.auer@germanistik.uni-freiburg.de.

Ethics statement

Ethical approval was waived by the Black Forest National Park. The study was conducted in accordance with the local legislation and institutional requirements. The participants provided written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

BL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was carried out as part of a cooperation of the Black Forest National Park, Department 3 (Dr. Kerstin Botsch), and the Chair of German Linguistics at the Albert-Ludwigs-University of Freiburg (Prof. Dr. Dr. h.c. Peter Auer), partly funded by the Black Forest National Park. I further acknowledge the support by the Open Access Publication Fund of the University of Freiburg.

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

Anna, M., and Pfeiffer, M. (2021). Die frühkindliche Verwendung des Exklamativformats oh+X: Noticings und Problemmitteilungen. *Gesprächsforschung* 22, 1–35.

Auer, P., and Laner, B. (2025). "Laughter and gaze among talkers on a walk" in Mobile Eyetracking: New avenues for the study of gaze in social interaction. eds. A. Stukenbrock and E. Zima (Amsterdam: John Benjamins).

Auer, P., Laner, B., Pfeiffer, M., and Botsch, K. (2024). "Noticing and assessing nature: a multimodal investigation of the format, perception imperative + exclamative' based on mobile eye-tracking data" in New perspectives in interactional linguistic research. eds. D. Barth-Weingarten and M. Selting (Amsterdam: Benjamins).

Auer, P., and Zima, E. (2021). On word searches, gaze, and co-participation. *Gesprächsforschung* 22, 390–425.

Botsch, K. (2021). "Fotografierpraktiken und die Herstellung von Raum im Go-Along-Interview" in Raum und Bild – Strategien visueller raumbezogener Forschung. eds. R. Kogler and J. Wintzer (Wiesbaden: Springer Spektrum).

Botsch, K., Auer, P., Laner, B., and Pfeiffer, M. (2025). Joint attention needs no language, but joint experience does. An eyetracking study on intersubjectivity and sensemaking in nature.

Burckhardt, L. (2006). Warum ist Landschaft schön? Die Spaziergangswissenschaft. *3rd* Edn. Berlin: Martin Schmitz Verlag.

Couper-Kuhlen, E.& and Selting, M. (2001). Introducing interactional linguistics. In: E. Couper-Kuhlen and M. Selting: Studies in interactional linguistics, 1–22, Amsterdam: John Benjamins.

Couper-Kuhlen, E., and Selting, M. (2018). Interactional linguistics: Studying language in social interaction. Cambridge: Cambridge University Press.

de Vries, C., Andries, F., and Meissl, K. (2024). Mocking enactments: a case study of multimodal stance-stacking. *Front. Psychol.* 15:1379593. doi: 10.3389/fpsyg.2024.1379593

Deppermann, A. (2021). "Imperative im Deutschen: Konstruktionen, Praktiken oder social action formats?" in Verfestigungen in der Interaktion. eds. B. Weidner, K. Katharina, I. Wolfgang and W. Lars (Berlin, Boston: De Gruyter), 195–230.

Deppermann, A., and Schmitt, R. (2007). Koordination: zur Begründung eines neuen Forschungsgegenstandes. In: Schmitt, Reinhold (Hg): Koordination: Analysen zur multimodalen Interaktion, vol. *38*. Tübingen: Narr, Studien zur deutschen Sprache, 15–54.

Du Bois, J. W. (2007). "The stance triangle" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam, Philadelphia: John Benjamins Publishing Company), 139–182.

Goffman, E. (1971). Relations in public: Microstudies of the public order. New York: Basic Books.

Goffman, E. (1978). Response cries. *Language* 54, 787–815. doi: 10.2307/413235

Goffman, E. (1981). Forms of talk. Philadelphia, PA: University of Pennsylvania Press.

Golato, A. (2002). German compliment responses. *J. Pragmat.* 34, 547–571. doi: 10.1016/S0378-2166(01)00040-6

Golato, A. (2012). German oh: marking an emotional change of state. *Res. Lang. Soc. Interact.* 45, 245–268. doi: 10.1080/08351813.2012.699253

Goodwin, C., and Goodwin, M. H. (1981). Conversational Organization: Interaction between Speakers and Hearers. London: Academic Press.

Goodwin, C., and Goodwin, M. H. (1986). Gesture and co-participation in the activity of searching for a word. *Semiotica* 62, 51–75.

Goodwin, C., and Goodwin, M. H. (1987). Concurrent operations on talk. Notes on the interactive Organization of Assessments. IPrA papers. *Pragmatics* 1, 1–54.

Goodwin, C., and Goodwin, M. H. (2004). "Participation" in A companion to linguistic anthropology. ed. A. Duranti (Malden, Mass: Blackwell), 222–244.

Goodwin, M., and Goodwin, C. (2012). "Car talk: integrating texts, bodies, and changing landscapes" in Semiotica, vol. *2012* (De Gruyter Mouton: Berlin/Boston), 257–286.

Günthner, S. (2017). "Diskursmarker in der Interaktion – Formen und Funktionen univerbierter guck mal- und weißt du-Konstruktionen" in Diskursmarker im Deutschen. Reflexionen und Analysen. eds. H. Blühdorn, A. Deppermann, H. Helmer and T. Spranz-Fogasy (Göttingen: Verlag für Gesprächsforschung), 103–130.

Haddington, P. (2006). The organization of gaze and assessments as resources for stance taking. *Text Talk* 26, 281–328. doi: 10.1515/TEXT.2006.012

Heritage, J. (2002). The limits of questioning: negative interrogatives and hostile question content. *J. Pragmat.* 34, 1427–1446. doi: 10.1016/S0378-2166(02)00072-3

Heritage, J. (2015). Well-prefaced turns in English conversation: a conversation analytic perspective. *J. Pragmat.* 88, 88–104. doi: 10.1016/j.pragma.2015.08.008

Heritage, J., and Raymond, C. (2005). The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. *Soc. Psychol. Q.* 68, 15–38. doi: 10.1177/019027250506800103

Jefferson, G. (1984). Notes on a systematic deployment of the acknowledgement tokens 'yeah' and 'mm hm'. *Papers Linguist.* 17, 197–216. doi: 10.1080/08351818409389201

Kääntä, L. (2014). From noticing to initiating correction: students' epistemic displays in instructional interaction. *J. Pragmat.* 66, 86–105. doi: 10.1016/j.pragma.2014.02.010

Keisanen, T. (2012). "Uh-oh, we were going there": environmentally occasioned noticings of trouble in in-car interaction. *Semiotica* 2012, 197–222. doi: 10.1515/sem-2012-0061

Kendon, A. (1990). Conducting interaction: Patterns of behavior in focused encounters. Cambridge: Cambridge University Press.

Kendrick, K. H., and Holler, J. (2017). Gaze direction signals response preference in conversation. *Res. Lang. Soc. Interact.* 50, 12–32. doi: 10.1080/08351813.2017.1262120

Kidwell, M. (2006). "Calm down!": the role of gaze in the interactional management of hysteria by the police. *Discourse Stud.* 8, 745–770. doi: 10.1177/1461445606069328

König, K. (2017). "Question tags als Diskursmarker? – Ansätze zu einer systematischen Beschreibung von ne im gesprochenen Deutsch" in Diskursmarker im Deutschen. Reflexionen und Analysen. eds. H. Blühdorn, A. Deppermann, H. Helmer and T. Spranz-Fogasy (Göttingen: Verlag für Gesprächsforschung), 233–258.

Krug, M. (2025). "Gaze aversion as a marker of disalignment in interactions." in Mobile Eyetracking: New avenues for the study of gaze in social interaction. eds. A. Stukenbrock and E. Zima (Amsterdam: John Benjamins).

Laanesoo, K., and Keevallik, L. (2017). Noticing breaches with non-polar interrogatives: Estonian *Kes* 'who' ascribing responsibility for problematic conduct. *Res. Lang. Soc. Interact.* 50, 286–306. doi: 10.1080/08351813.2017.1340721

Laner, B. (2022). 'Guck mal der Baum' – Zur Verwendung von Wahrnehmungsimperativen mit und ohne mal. *Gesprächsforschung* 23, 1–35.

Lehmann, A. (1999). Von Menschen und Bäumen. Die Deutschen und ihr Wald. *1st* Edn. Reinbeck bei Hamburg: Rowohlt.

Merlino, S., and Mondada, L. (2019). Crossing the street: how pedestrians interact with cars. *Lang. Commun.* 65, 131–147. doi: 10.1016/j.langcom.2018.04.004

Mondada, L. (2014). Bodies in action: multimodal analysis of walking and talking. *Lang Dial* 4, 357–403. doi: 10.1075/ld.4.3.02mon

Mondada, L. (2017). Walking and talking together: questions/answers and mobile participation in guided visits. *Soc. Sci. Inf.* 56, 220–253. doi: 10.1177/0539018417694777

Pekarek Doehler, S., Polak-Yitzhaki, H., Li, X., Stoenica, I. M., Havlík, M., and Keevallik, L. (2021). Multimodal assemblies for prefacing a Dispreferred response: a cross-linguistic analysis. *Front. Psychol.* 12:689275. doi: 10.3389/fpsyg.2021.689275

Pfeiffer, M. (2017). The deictic dimension of exclamations: on the use of wh-exclamatives in German face-to-face interaction. *Revue de Sémantique et Pragmatique* 40, 35–57. doi: 10.4000/rsp.288

Pfeiffer, M., and Anna, M. (2021). Recruiting assistance in early childhood: longitudinal changes in the use of 'oh+X' as a way of reporting trouble in German. *Res. Lang. Soc. Interact.* 54, 142–162. doi: 10.1080/08351813.2021.1899708

Pillet-Shore, D. (2020). When to make the sensory social: registering in face-to-face openings. *Symb. Interact.* 44, 10–39. doi: 10.1002/symb.481

Pomerantz, A. (1984). "Agreeing and disagreeing with assessments: some features of preferred/Dispreferred turn shapes" in Structures of social action: Studies in conversation analysis. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 57–101.

Rauniomaa, M., Keisanen, T., and Siitonen, P. (2019). Shared understandings of the human–nature relationship in encounters with small wildlife. Amsterdam: John Benjamins.

Robinson, J. D. (2020). One type of polar, information-seeking question and its stance of probability: implications for the preference for agreement. *Res. Lang. Soc. Interact.* 53, 425–442. doi: 10.1080/08351813.2020.1826759

Rühlemann, C., Gee, M., and Ptak, A. (2019). Alternating gaze in multi-party storytelling. *J. Pragmatics* 149, 91–113. doi: 10.1016/j.pragma.2019.06.001

Ryave, A. L., and Schenkein, J. N. (1974). Notes on the art of walking. In: R. Turner (Hg.), Ethnomethodology: Selected readings. Harmondsworth: Penguin Education, 265–274.

Sacks, H. (1992). Lectures on Conversation, vol. *II*. Oxford: Blackwell.

Sacks, H. (1995). Lectures on conversation. Oxford: Basil Blackwell.

Schegloff, E. A. (1986). The routine as achievement. *Hum. Stud.* 9, 111–151. doi: 10.1007/BF00148124

Schegloff, E. A. (1988). "Goffman and the analysis of conversation" in Erving Goffman: Exploring the interaction order. eds. P. Drew and A. J. Wootton (Cambridge: Polity Press), 89–135.

Schegloff, E. A. (2007). Sequence organization in interaction: A primer in conversation analysis. Cambridge: Cambridge University Press.

Schegloff, E. A., and Sacks, H. (1973). Opening up closings. *Semiotica* 8, 289–327. doi: 10.1515/semi.1973.8.4.289

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung* 10, 353–402.

Sidnell, J., and Stivers, T. (2013). The handbook of conversation analysis. Oxford: Wiley-Blackwell.

Siromaa, M., and Rauniomaa, M. (2021). "Stance and evaluation" in The Cambridge handbook of Sociopragmatics. Cambridge handbooks in language and linguistics. eds. M. Haugh, D. Z. Kádár and M. Terkourafi, vol. *2021* (Cambridge: Cambridge University Press), 95–116.

Steensig, J. (2019). "Conversation analysis and affiliation and alignment" in The encyclopedia of applied linguistics (Atlanta, GA: American Cancer Society), 1–6.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: when nodding is a token of affiliation. *Res. Lang. Soc. Interact.* 41, 31–57. doi: 10.1080/08351810701691123

Stivers, T., and Rossano, F. (2010). Mobilizing Response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Stukenbrock, A. (2015). Deixis in der face-to-face-Interaktion, 47. Berlin/Boston: de Gruyter.

Stukenbrock, A., and Dao, A. N. (2019). Joint attention in passing: what dual mobile eye tracking reveals about gaze in coordinating embodied activities at a market. In: E. Reber, Gerhardt, C. (eds.): Embodied activities in face-to-face and mediated settings, Palgrave Macmillan, Cham, 177–213.

Szymanski, M. H. (1999). Re-engaging and Disengaging Talk in Activity. *Lang. Soc.* 28, 1–23.

Weiß, C. (2020). Blick und Turn-Taking in Face-to-Face-Interaktionen. Multimodale interaktionsanalysen triadischeer Gesprächssituationen mit Hilfe von Eye-Tracking. Dissertation, Universität Freiburg. Verlag für Gesprächsforschung.

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 24 February 2025
DOI 10.3389/fpsyg.2024.1479825

Check for updates

OPEN ACCESS

EDITED BY
Elisabeth Zima,
University of Freiburg, Germany

REVIEWED BY
Barbara Lewandowska-Tomaszczyk,
State University of Applied Sciences in Konin, Poland
Paul Sambre,
KU Leuven, Belgium

*CORRESPONDENCE
Barbara Dancygier
barbara.dancygier@ubc.ca

†These authors have contributed equally to this work and share first authorship

RECEIVED 12 August 2024
ACCEPTED 05 November 2024
PUBLISHED 24 February 2025

CITATION
Dancygier B and Vandelanotte L (2025) Embodiment and simulated interaction in online stance expression.
Front. Psychol. 15:1479825.
doi: 10.3389/fpsyg.2024.1479825

COPYRIGHT
© 2025 Dancygier and Vandelanotte. 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.

Embodiment and simulated interaction in online stance expression

Barbara Dancygier[1*†] and Lieven Vandelanotte[2,3†]

[1]Department of English Language and Literatures, The University of British Columbia, Vancouver, BC, Canada, [2]Namur Institute of Language, Text and Transmediality, Université de Namur, Namur, Belgium, [3]Research Group Functional and Cognitive Linguistics, Department of Linguistics, KU Leuven, Leuven, Belgium

The expression and exchange of stance drives much social media discourse, including internet memes. We demonstrate how, even in the absence of actual face-to-face communication, online discourse and memes rely on the dynamics of embodiment and dialogue in comparable ways, while also developing specific constructional forms for this with no direct face-to-face equivalent. We introduce the notion of simulated interaction to refer to the combinations of embodied expression, images, and the structures of (apparent) quotation and dialogue allowing online communicators to vividly represent experience and signal stance.

KEYWORDS

constructional forms, dialogue, images, internet memes, multimodality, simulated interaction, social media discourse, stance

1 Introduction and approach taken

Online discourse as it unfolds on various social media platforms is rife with expressions and exchanges of stance. It is no exaggeration to say stance exchange is the lifeblood of a platform such as X/Twitter: expressing viewpoints which can then be 'liked', shared, commented on, or quoted with or without further comment is what continually fills up the so-called 'timeline' people read, refresh and respond to (e.g., Wikström, 2019; Vandelanotte, 2020). Similarly, internet memes are driven by internet users' need to make light of, critique or ironize all sorts of daily frustrations, large or small, and their popularity increasingly molds discourses online as well as offline (Dancygier and Vandelanotte, 2017b; Dancygier and Vandelanotte, 2025).

Internet memes in particular have been the focus of much popular interest as well as scholarly research into how they can be defined, how they emerge and how they contribute to public discourse (e.g., Jenkins, 2014; Shifman, 2014; Wiggins and Bowers, 2015; Milner, 2016; Wiggins, 2019). The foundation laid by this line of research has clearly demonstrated the degree to which the notion of memes itself has evolved since the term's initial coining by Dawkins (1976), where memes were defined as units of cultural replication, on the model of 'genes' as 'selfish' and virulent units of genetic replication; examples include, e.g., catchphrases, tunes, fashions, architectural styles and the like. Internet memes are quite different in clearly involving the active agency and creativity of meme communicators, who are not more or less passive 'hosts' but who deliberately devise textual and visual modifications and remixes of established meme patterns in order to target further areas of experience they want to express stance toward.

Internet memes and other forms of online discourse have generated considerable study of political meanings and social ramifications (e.g., Milner, 2016; Ross and Rivers, 2017; Denisova, 2019; Paz et al., 2021; Zappavigna and Logi, 2024). Also, studies of pragmatic aspects of such discourse types have touched on questions such as common ground, intertextuality and humorous incongruity (e.g., Yus, 2018; Wikström, 2019; Xie, 2022; Attardo, 2023). The strand of research

we have broadly approached these discourse types from, and on which we build here, is one inspired by the tenets of construction grammar and cognitive linguistics, as represented in, e.g., Dancygier and Vandelanotte, (2017b); Dancygier (2017); Lou (2017); Zenner and Geeraerts (2018); Bülow et al. (2018); Piata (2020); Vandelanotte (2021); Kang et al. (2023). This line of work treats internet memes as multimodal constructions, i.e., pairings of form and meaning. Specific memetic forms are governed by rules which can be described in a 'grammar' of memes, while the meaning of a meme emerges from integrating the meaning of the memetic template with the situations and frames evoked, with important roles in this meaning emergence played by figuration and viewpoint. Dancygier and Vandelanotte (2025) propose an overview of the patterns of meaning typically involved, across image and text.

The role of stance in internet memes and social media discourse more generally has been highlighted (e.g., Shifman, 2014; Dancygier and Vandelanotte, 2017b; Droz-dit-Busset, 2022), but we propose to analyze specific patterns of interactions used in internet memes effectively for the purposes of stance exchange. Our understanding of stance is informed by the work of Du Bois (2007), who defines stance as "a public act by a social actor, achieved dialogically through overt communicative means (language, gesture, and other symbolic forms), of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field." Among the aspects of this definition which we want to highlight are its multimodal orientation, not restricting itself to the verbal stream in communication (*cf.* Andries et al., 2023), and its posit of stance as a 'tri-act': "I evaluate something, and thereby position myself, and thereby align with you." Du Bois also rightly stresses the frequent resonance of "the current stance act … with a stance taken in prior discourse"—a scenario of which, we would argue, much online discourse exchanges present a heightened case—and introduces the terms of "stance lead" and "stance follow" to refer to an initial stance and a response to it. Dancygier (2012), in her analysis of negation of stance verbs, has introduced the notion of stance-stacking to characterize constructions which accommodate multiple stances being stacked one upon the other. We believe many examples of internet memes incorporate such complex stance configurations. More specifically, they often involve forms, such as facial expressions, body postures and scenes of interaction and dialogue between agents, that lend themselves to emotional interpretation, and are often used to vent about life's daily smaller or larger frustrations, or about people's amusing or annoying foibles. Our focus is thus on the expression of emotional or affective stances (Ochs, 2006; Goodwin et al., 2012; Couper-Kuhlen and Selting, 2017), where we concur with Goodwin et al. who argue that emotion is a "contextualized, multiparty, multimodal process" (2012: 18) which "emerges in unfolding interaction" (2012: 24).[1]

2 Materials and research questions

The present article is an offshoot of the larger project reported on in Dancygier and Vandelanotte (2025), which focuses on proposing a 'grammar of memes', demonstrating the emergence of constructions in which images become structural components, while making language forms adjust to the emergence of multimodal rules. We base this project, and also the current article, on a large and growing collection of internet memes, relying on a combination of meme collecting sites (such as Knowyourmeme.com), internet searches, and our own observations on social media platforms such as X/Twitter, Facebook and Reddit.[2] Our aim in data collecting is to discover the range and classes of internet meme types that can usefully be distinguished along major parameters, such as the involvement of *ad hoc* vs. more entrenched images, various kinds of grid-like arrangements of the space of the meme, use of fictively quoted clauses and lines of dialogue, and the different uses and arrangements of text (e.g., using lines of top text and bottom text within the space taken up by the image, vs. using text in a blank space above the image, vs. using single-word or phrasal labels on top of individual components within the image). The aim is thus not to capture the minutiae of pop culture references in memes (which existing online resources do very well), nor each and every new minor variation in meaning expression, but rather in identifying conventionalizing or conventional form-meaning patterns in the data.

In this article, we re-examine some of our data, and include new cases, to elucidate two main research questions. Both of these questions find their origin in the specific type of 'multimodality' involved in the kinds of image-text combinations we will consider. Where the special issue's main focus is on embodied interaction in face-to-face settings, or in mediated, virtual forms of such settings, we look at online discourse that is not face-to-face. Relatedly, our interest here lies with combinations of image and text—different semiotic resources accessed visually—rather than with combinations of language and embodied behavior such as gesture, gaze and posture. While we recognize these frameworks are different and are thus sometimes referred to using different terminology (such as polysemioticity vs. multimodality), we have always stressed the potential for fruitful crossfertilization between the two approaches (Dancygier and Vandelanotte, 2017a, Vandelanotte and Dancygier, 2017). In this article specifically, we therefore turn to two questions which a reconsideration of our recent data sets (Dancygier and Vandelanotte, 2025) have thrown up: (i) what role do visual depictions of forms of embodied behavior (including facial expressions, gestures, postures) play in internet memes, and how; and (ii) what role do

1 While we recognize the distinctions drawn in psychology between affect and emotion (as reported, for instance, in Barrett (2017): Chapter 4), we accept that the term 'affective stance' has gained wider currency in linguistics-oriented work, and note that the term is used interchangeably with 'emotion as stance' in Goodwin et al. (2012). We think our examples of stance expression are more complex and fine-grained, and more situation-bound, than the kind of basic, ever-present sense of affect (in its dimensions of valence and arousal) described by Barrett (2017), and so seem to have more to do with emotion.

2 An internet meme generally has many creative mothers and fathers. Even in a case where one might know specifically who made one particular meme object, it will rely on pre-existing and pre-circulating visual material as well as on the pattern of variation communally established; more often, it is simply not, or no longer, possible to verify which internet account (let alone which 'biographical' person) posted a particular iteration or innovation. For a thoughtful reflection on this, we refer to Milner's (2016, pp. 221–232) appendix on methods and ethics. Wherever possible, in Section 3.5 in particular, where we can source an example to a specific post on X/Twitter, we have provided the source information.

'pretend-conversations', sometimes extending over multiple lines of exchange, and often including images as 'turns' in an exchange, play in memes and other social media discourse? While the artifacts we study are very different from face-to-face settings ('real' or virtual) in which discourse develops over time, our concluding section reflects on the relevance of including such forms of multimodality in investigations of the embodied interactions they so succinctly reflect back at online communicators. In addition, we contribute to the theorizing, in cognitive linguistics, of Pascual's key notion of fictive interaction (Pascual, 2002, 2014; Pascual and Sandler, 2016), by proposing the term 'simulated interaction' to capture the more complex forms our examples bring to light. 'Simulated interaction' is, in the approach we propose, a communicative mode which builds on familiar conversational patterns and mechanisms of 'multimodality in interaction' (Feyaerts et al., 2017a) to give a salient form of expression to evaluative and emotional stances. We rely on Pascual's idea of the 'conversation frame', while showing how 'simulated interaction' develops its own formal parameters and constructional forms, and how these emergent mechanisms serve the purposes of stance expression in online discourse. We rely on internet memes as our primary online discourse genre, because their form and meaning provide a succinct but compelling demonstration of the forms and functions of simulated dialogic patterns.

FIGURE 1
One Does Not Simply meme. Reproduced from Imgflip.com.

FIGURE 2
Good Girl Gina meme. Reproduced from Imgflip.com.

3 Analysis and discussion

3.1 Depicted bodies in internet memes constructing emotional stance

Internet memes are typically combinations of images and text. A more elaborate analysis of the role these two semiotic modes play in memes is beyond the scope of this paper, and so we will focus on some limited observations regarding the role memetic images play in constructing stance.

Memetic images are essentially of two kinds—entrenched or non-entrenched, but most (if not all) of them use representations of body postures and facial expressions. The non-entrenched images are often random selections made by the Meme Maker, guided by the easily recognized expressivity of the posture or face represented, while the entrenched ones play a unique role of simulating a personality or event type.

Classic examples of so-called image macro memes, where images are entrenched, combine a recurring image, emblematic of the meme pattern, with text neatly divided across top and bottom areas of the space taken up by this recurrent image. Two very basic examples which we discussed in previous work (Dancygier and Vandelanotte, 2017b) can be relied on here to illustrate the case; both reflect sexist stereotypes which exist around women and arguments. The example of the One Does Not Simply meme, shown in Figure 1, suggests it is futile, to the point of impossibility, to try to win an argument with a woman. In one example of the so-called Good Girl Gina meme (Figure 2), the image of the meme's stock character (a smiling woman looking very happily and confidently into the camera) is accompanied by the lines of text "gets mad of you" (as top text, i.e., shown near the upper edge of the image) and "tells you why" (as bottom text, shown near the lower edge). The example suggests that it's only a truly exceptionally "good girl" like "Gina" who will tell you why she got mad at you (implying that most women, when they are angry, stubbornly expect you to just know or guess why).

Both these examples depict faces of Meme Characters—Boromir, a character from the film *The Lord of the Rings* in the One Does Not Simply example, and the stock character of Good Girl Gina, taken to represent women who display highly considerate and virtuous behavior. The memes are not 'about' these characters as actual 'referents'; rather, they are used to communicate generally recognizable scenarios and the Meme Maker's stances about these, within an intersubjective context where it is expected that the meme communicator's primary online audience can relate to and perhaps share the attitudes expressed. As we have argued in other work (Dancygier and Vandelanotte, 2017b; Dancygier and Vandelanotte, 2025), image macro memes perform two related tasks—the meme template establishes a category of stance (e.g., the belief that some goals cannot be achieved, in the One Does Not Simply case), while every next meme building on the template represents a new instance of the specific stance-evaluated behavior (e.g., every next example of what Good Girl Gina would do in a situation is evaluated as welcome and generally positive).

While the embodiment on display is necessarily limited and stilted—with only a still image available—even in such a simple case, the facial expressions, postures and gestures are very important in construing the overall meanings. In the One Does Not Simply example, the combination of the still from the film and the signature line "one does not simply"—originally used to talk about the

dangerous and near-impossible task of 'walking into Mordor'—together build the category of unachievable endeavors, to which meme communicators add new examples with each new bottom line of text. The strained look of concentration on Boromir's face, and the 'point-making', focused accompanying hand gesture, are part of what lends the original film scene its intensity and significance. Part of the effect of the meme then lies in the almost anticlimactic application of this portentous context to everyday gripes and frustrations which are hardly of the life-threatening variety, resulting in a mildly mocking, ironic overall viewpoint of venting about life's smaller difficulties, perhaps even hinting at an awareness of the (sexist) stereotypes involved.

In the Good Girl Gina example, the smiling face of Gina, resting on the right hand and looking happily and confidently into the lens, is meant to call up the personality type of the stock character of the 'good girl' (part of a cast of Meme Characters, including Scumbag Steve and Good Guy Greg, among others). Proficient meme communicators know, when they see the image macro of Good Girl Gina, that a generic "when/if X-then Y" 'predictive' reasoning (Dancygier, 1998) will be presented, giving another instance of considerate female behavior. In the example here, we could reconstruct the kind of meaning prompted for, with the implied sexist nod to the meme audience, as something like this: 'when a considerate woman gets mad at you, then (unlike what you and I usually experience at the hands of women) she will explain the reasons for her anger so you can respond appropriately'. The integration of image and text is so tight in this meme construction that it allows suppression of the subject argument, normally required in English grammar, given that in this genre, it is provided visually, via the image macro: *when (Good Girl Gina) gets mad at you*. This can be seen as a multimodal application of Ruppenhofer and Michaelis' (2010) notion of genre-based argument omission, which they discuss in relation to the language of, for instance, diaries or recipes, where the subject or the object is routinely left out thanks to the conventions of the specific genre (e.g., *Went to the cinema* for a diary entry; *place on the stove and bring to a simmer* for a recipe).

More generally, we would argue that for meme proficient communicators—people who are familiar with how memes construct meaning—the meanings of memes are as tightly constrained as those of ordinary linguistic constructions: for firmly entrenched meme constructions such as One Does Not Simply or Good Girl Gina, the memetic conventions and the form-meaning patterning very strongly cue the intended meanings. Of course, there may be occasional Meme Viewers who are not familiar with memetic grammar in general, or with a particular (perhaps only just emerging) memetic template, especially when it relies on culturally rich frames which the viewer may not be familiar with. For entrenched cases, though, knowledge of such frames (for instance, that of Boromir and *The Lord of the Rings*) is arguably less important: a communicator who is 'very online' will learn to understand and use the meme even without having seen the film scene, even if perhaps they may draw less aesthetic or humoristic enjoyment from the experience compared to Tolkien aficionados.

It is important to add at this point that the still images provided in internet memes are not always unambiguously interpretable in emotional terms. We may think we can easily 'read faces', and body language more broadly, but usually we have very rich contextual cues, missing from the still image. The more general point still, made by Barrett (2017), is that facial expressions do not automatically and unambiguously represent emotions. Thus in a case like Good Girl Gina, we construe our understanding of the Meme Character's facial expression to fit the meaning of the meme. This is perhaps more striking still in some of the other stock characters that are part of the series: Scumbag Steve, for instance, does not look particularly awful or unpleasant (his expression could just as well be construed as nonplussed, clumsy or even shy), or the (less commonly used) character of Good Guy Greg does not show a face which we automatically take to be that of someone you can definitely trust and rely on to do the right thing. In the broader history of internet memes, one well-known example has come to be known as 'Sudden Clarity Clarence', described on Knowyourmeme.com as featuring a "young man at a party staring into the distance as if he is experiencing an epiphany." The clue is in the "as if," which illustrates Barrett's point about interpretation; we do not actually know if there was indeed 'sudden clarity' or an epiphany taking place, but this is the meaning construed in the resulting internet meme.

Given these observations, it stands to reason that image, text, and memetic grammar cooperate in determining our interpretations of embodied features in the images featured in internet memes. As well, we assume the degree to which a given image is open to varying interpretations and emotional alignments can vary. Let us first briefly highlight a few examples—across different types of meme patterns—that are quite strongly premised on cueing for clear categories of emotional stance—essentially likes and dislikes—and where the main embodied features contributing to this stance construal are in Meme Characters' bodies. Two that feature image macros—where the images are a constant recurring feature across the instances of the meme—are the Drake meme and the Two Guys on a Bus meme. The former features two contrasting stills from a music video by hiphop artist Drake, arranged in a grid with the two images on the left, and matching words (or sometimes pictures again) on the right (for an example, see Figure 3). The embodied postures in the two Drake images involve a very pronounced opposition, one showing dismissal and dislike (where Drake turns away from what is presented to him and displays an open-palm-out 'blocking' or 'stopping' hand gesture), the other suggesting acceptance and appreciation (Drake standing physically close to what is presented to him, sporting a satisfied grin and pointing the index finger toward the presented text, in apparent approval). The simultaneous presence of the two recurrent images helps Meme Viewers confirm that construal of contrasting stances is involved—firm rejection and blissful acceptance. An example where a single image macro incorporates two opposing stances is that of the Two Guys on a Bus meme, exemplified in Figure 4. This is a so-called labeling meme (e.g., Vandelanotte, 2021; Dancygier and Vandelanotte, 2025), in which parts of a single scene presented in an image are overlaid with words or phrases that do not identify those parts in the depicted scene, but rather call up an entirely different frame (the best-known example being the Distracted Boyfriend meme, applying the shift in romantic attention to other shifts in attention, preference or allegiance; see, e.g., Walker, 2023). In the Two Guys on a Bus meme, a desperately sad looking man on the left hand side of a bus is looking at a dark rock wall, whereas a happy looking man on the right hand side enjoys (and apparently photographs) a sunny view of a mountainous landscape; the labeling then applies this, typically, to contrasting stance-marked situations, e.g., 'does not smoke cigarettes' vs. 'smokes cigarettes'. Importantly, the 'good' versus 'bad' option is

FIGURE 3
Drake meme. Reproduced from Imgflip.com.

FIGURE 4
Two Guys on a Bus meme. Reproduced from Imgflip.com.

represented in a rather complex way. There is the view outside (beautiful or boring), there is the idea of smoking (bad for your health) versus non-smoking (good), and the Meme Characters' faces (happy/sad). But the way these stance contrasts align is complex: nice-view-and-good-mood are aligned with smoking (which should be marked as 'bad'), while boring-view-and-bad-mood are aligned with non-smoking. It is this inconsistency that leads the Meme Viewer to the conclusion that the Meme Maker considers smoking to be good (something that brings pleasure)—and that way both sets of stances are coherently aligned. In other words, the generic stance toward smoking is overridden via meme-internal stance alignments, which are a more important determinant of stance than independent knowledge.

Our examples so far have featured entrenched images—so-called image macros that are reused again and again, but with altered text. Non-entrenched images, that are not themselves fixed, can likewise serve as strong prompts for the construal of emotional stance. So-called *when*-memes, analyzed extensively in Lou (2017, 2021) as a case of multimodal simile, often present such cases. In a *when*-meme, there is a single *when*-clause, not completed textually, but instead followed by a suitable '*ad hoc*' image to fill the missing 'slot' (the *then*-part of a predictive construction). The image is unrelated to the scenario described in the *when*-clause, except in one important respect: it shows a response or situation which *feels* the same (hence Lou's proposal to treat *when*-memes as cases of 'simile'). One example is given in Figure 5, where the recognizable situation of a boss asking an employee to do one final (possibly complex) bit of work just before work time is scheduled to end, makes the employee *feel* the same kind of annoyance or disappointment as could be seen in the facial expression of the actor Leonardo DiCaprio (not, in fact, being told to work late, but appearing at some red carpet event). Even if the disappointed face may seem to be just an expression of an individual's emotion, we agree with Goodwin et al. (2012, p. 17) that "the scope of an emotion is *not* restricted to the individual who displays it," and that "emotions constitute public forms of action." Even though we only see one person depicted, that person's emotion still communicates something in an interactive setting. The second person pronoun *you* typically featured in *when*-memes is generic, rather than deictic: there is no specific deictic ground with a specific addressee (Dancygier, 2021), and this evokes the assumption that the Meme Viewer can share the Meme Maker's stance. The meaning of the meme overall is paraphrasable as a similative statement such as, *When you are about to leave work and the boss says "before you go…," you feel like Di Caprio felt when not being awarded an Oscar yet again*.

Interestingly, several of Lou's (2017, 2021) examples, and many other *when*-memes beyond, involve not humans, but animals—something also seen in purely textual simile (e.g., *as proud as a peacock*, *cf.* Veale, 2012). For instance, a photo of a koala attaching itself firmly to a person's ankle completes the *when*-clause *when you are at a party full of people you do not know so you u stay with ur friend the whole time* (Lou, 2017, p. 121). Similarly, many *when*-memes use existing paintings (*cf.* Piata, 2020), as with the example of Degas' famous absinthe drinker, preceded in a *when*-meme by the clause *when you are on your lunch break and consider not going back*. Thus in *when*-memes in general the images represent salient postures and facial expressions which match the experience described in the *when*-clause. The focus on embodied representations of emotions and experiences makes *when*-memes interesting cases of exploitation of visual representations of embodied behavior, especially because the person actually feeling something (disappointment or social awkwardness), and using the meme to express this feeling (i.e., the Meme Maker), is not represented in the image at all.

Such *when*-memes are thus quite directly about the similarity of stance evoked by two different situations: one described textually, the other, unrelated, prompted for in the image. More generally, the examples we have seen in this section show how images of faces, gestures and postures help prompt Meme Viewers to simulate interactional stance meanings, applying them to the situations and events described in the meme text. They achieve this despite the fact that in these examples, the meme does not show interaction between people, leaving a fuller interaction involving other participants to be inferred. In the next section, we turn to embodied interactions

FIGURE 5
When-meme depicting a disappointed face. Reproduced from Boredpanda.com.

FIGURE 6
Distracted Boyfriend meme. Original photo by Antonio Guillem, licensed via Shutterstock.

between multiple depicted bodies, multiplying the opportunities for stance expression.

3.2 Multiple bodies, multiple stances

We now turn to a different form of stance complexity: as soon as more than one body is represented in the image featured in a meme, an interpersonal dynamic can be activated, expressed in embodiment terms but also in concrete actions and interactions between the people depicted. The emergence of interpersonal stance configurations is not automatically evoked, as we could see in the Two Guys on a Bus meme, where the stances represented are relevant to the Meme Maker, not to a scenario in which the Meme Characters interact.

The popular Distracted Boyfriend meme, however, is a different case. Each individual Meme Character in the image macro has relevant facial features: suggesting admiration and heightened interest on the part of the central male character, confidence to the point of self-satisfaction in the passing girl in red, and annoyance and anger in the newly ignored girl in blue, holding hands with the male character. All of these interpretations only really make sense combined with the movements and postures shown, in particular the man's turning to keep looking, over his shoulder, at the admired girl in red passing by. These emotional dynamics, played out in human relationships we can easily recognize, then become applied to an unrelated frame through the application of textual labels, as with the dieting frame in the example in Figure 6, where the lure of pizza proves too strong for someone supposedly on a diet. Overall, the meme provides an ironic comment on all too human foibles—preferring the new attractive thing over something that we already have and that is good for us. Another labeling meme that shows an interaction—in this case with very strong force dynamic impact (Talmy, 1988) resulting from a physical altercation—is the Will Smith Slapping Chris Rock meme, based on a still from an incident during the 2022 Oscars ceremony in which Smith slapped Rock in response to a joke the latter made about the former's wife. In one example, for example, the attacker (Smith) is labeled as "Monday," and the undergoer (Rock) as "me trying to enjoy the weekend," applying the strongly negative, punishment-exacting stance to the effect the start of the work week has on people coming out of a weekend.

In addition to depictions of people (or, as we discussed in Section 3.1, animals), memes may also rely on cartoon depictions. An interesting example which shows important embodied emotional stance is that of the Drowning High Five meme, based on a cartoon by artist Gudim. The event structure here is particularly rich, as four scenes develop across the four panels of the grid (see Figure 7): (1) a hand outstretched from a large expanse of water (suggesting "not waving, but drowning," to quote the famous Stevie Smith poem); (2) another hand approaching from the upper left hand corner, suggesting an approaching offer of help, and thus a positive, hopeful emotional stance; (3) a high five gesture—itself expressive of very strong positive stance, usually congratulatory—being performed by the approaching hand; (4) back to the original, single hand sticking out of the water, but more deeply submerged than in the first frame, surrounded by bubbles suggesting further immersion into the deep waters—signifying that the hope felt earlier turned out to be but false hope, and no actual help was offered. A very large portion of existing Drowning High Five memes, including the example in Figure 7, apply the original scene of physical distress (drowning) to the domain of mental distress, as a way of commenting on well-meaning but ultimately useless advice around depression (represented by snippets of mental health advice being quoted, such as "hang in there," or "you have such a good life compared to some, just be happy").

The simulated expressivity in the examples above is closely correlated with some naturally recognizable aspects of facial expression and/or body posture. Throughout the discussion above we focused on aspects of embodied expression or behavior which are read as expressions of stance. There are differences in pathways of stance simulation across the examples we presented (and, naturally, many other examples), such that *when*-memes (Figure 5) may signal stance through easily recognizable facial expressions, while image macro memes like Good Girl Gina rely on the entire template to present the behavior described as pleasant and praiseworthy. Specific

FIGURE 7
Drowning High Five meme. Reproduced from Reddit.com user BasicSadBish, 27 March 2019.

cases may vary, but it seems justified to postulate a formal dimension present in many memes—facial expression and body posture as tokens of stance.

However, while memes are in general a stance-focused and image-based genre of internet discourse, many of them also include token snippets of what looks like conversation. In Figure 5, for one, the disappointed facial expression is triggered by a quoted phrase "before you go," which forces the Meme Character to abandon the plan to leave work for the day. The Drowning High Five grid meme shows quoted discourse as well, while also creating a story—a sequence of events which starts with a plea for help and ends in drowning. These two features—apparent discourse snippets prompting a narrative through correlation with an image or images—are a very salient aspect of memetic form and meaning. In Section 3.3 we will set the agenda for exploring memetic narratives and the role of apparent quotations in them, also showing how such discourse structures depend crucially on memetic images. Sections 3.4 and 3.5 will then offer a broader discussion of relevant examples and memetic meaning-making strategies, focusing on verbal interactions between speakers which may be explicitly depicted (Section 3.4) or, across a variety of formats, not depicted (Section 3.5).

3.3 Narratives, discourse snippets and memetic simulation

Our first example is the It Will Be Fun, They Said meme, in Figure 8. In general, this meme type condenses a story sequence across two lines of text and a picture: 'advice – acting on the advice – distressing outcomes that contradict the original advice'. The 'advice' part is given in the piece of discourse presented in the top line of text: "Go to grad school, they said," complemented by the positive stance expressed by the same unidentified "them" in the bottom line of text: "It will be fun, they said." Combining this with the negative emotional stance prompted by the picture of the inconsolably crying man leads viewers to pragmatically infer a scenario where the advice was followed, but led to bitter disappointment. Note that the 'lower' positive and negative emotional stances of promised fun and actual distress feed into an overall Discourse Viewpoint (Dancygier and Vandelanotte, 2016), which takes an ironic stance toward the whole narrative 'reversal of stance' sequence.

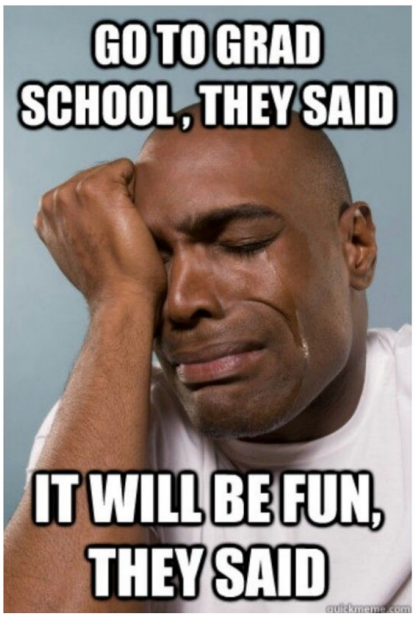


FIGURE 8
It Will Be Fun, They Said meme. Reproduced from Quickmeme.com.

The meme in Figure 8 is a good example of the many ways in which memes construct stance: by telling a brief but emotionally loaded story, by using representations of conversational discourse as tokens of narrative events, and by profiling the image as a representation of the final event in the story. In Figure 8 the language of the meme is a more or less standard form of Direct Speech, where the utterances are made to look as if they were quoted verbatim, and the reporting clause (*they said*) completes the construction. We are referring to this example here as a predictable pattern of a narrative which includes dialogic pieces. In our further discussion, we will show how memes construct stories which build the dialogic structure by inserting images in their place. In this case, though, the image represents the concluding event in the sequence and supports the simulation of emotional meanings which constitute the core of the meme.

In what follows we will highlight the important role of simulation in the emergence of emotional meaning in memes such as the ones discussed here (*cf.* Sweetser, 2012; Bergen, 2012; Feyaerts et al., 2017b). In understanding linguistic and visual inputs, our brains "run embodied simulations" (Bergen, 2012, p. 195); in Barrett's succinct summation, "[s]imulations are your brain's guesses of what's happening in the world" (Barrett, 2017: 27). Our capacity to recognize and relate to depictions of embodied experiences allows us to 'fill in' and also simulate the impact of the partial visual inputs the memes provide us with: we 'feel' (at least in an attenuated sense, by simulation), and can empathize with, the frustration felt by the girl in blue in the Distracted Boyfriend meme, for instance, or what it feels like to cry uncontrollably as in the It Will Be Fun, They Said meme; likewise, we can simulate, perhaps even with some sense of dread or of flinching respectively, what drowning feels like (in the Drowning High Five meme), or what a slap to the face is like (in the

Will Smith Slapping Chris Rock meme). Even if animals, or cartoon characters, or film characters, rather than 'ordinary' human characters are depicted, as soon as *bodies* are depicted, we are primed and ready to be attuned to feelings felt and emotions experienced.

The simulations are further supported by the use of memetic quotation—i.e. not the most standard type of quotation taking place in fully detailed deictic grounds, with fully identified speaker, addressee, time and place coordinates (*cf.* Dancygier, 2021), but rather snippets of discourse being used to quickly evoke frames and attitudes. We zoom in on discourse exchanges of such a 'fictive' kind (Pascual, 2002, 2014; Pascual and Sandler, 2016) in the remainder of this paper, first in examples of internet memes showing Meme Characters engaging in verbal exchanges (section 3.4) and then in memes and X/Twittter discourse making broader innovative use of longer exchanges in a kind of faux dialogue format identifying speech participants not visually, but by means of noun phrases ("me," "partisan Twitter," etc.) (section 3.5). We will also discuss the role of images in these dialogic formats and argue that the complexity of such usage calls for a more specific investigation of how the unusual discourse and visual forms yield the representation of stance. We will refer to such instances as 'simulated interaction'.

3.4 Depicted bodies interacting verbally in dialogue memes

In research on quoted speech generally, the view that quotations are demonstrations or 'depictions' (Clark and Gerrig 1990; Clark, 2016) has been highly influential—and suits many memetic quotation examples well: rather than having 'actually' been said by some identifiable speaker to a specific addressee, memetic quotes are used to quickly point up attitudes and stances, which can then be responded to by other components of the same meme. A typical, older example is that of the Said No One Ever meme (Dancygier and Vandelanotte, 2016), for instance in an example like "I love your Crocs, said no one ever": the postposed speech clause "said no one ever" effectively reverses the viewpoint initially expressed. Apparently, in the Meme Maker's view, the idea that anyone could love anyone's Crocs is so laughable that it could not be expressed by anyone ever. Our earlier It Will Be Fun, They Said example (Figure 8) shows another typical kind of stance reversal, as do many examples of Be Like memes (Vandelanotte, 2019; Dancygier and Vandelanotte, 2025).

Particularly where more speakers are involved in a non-actual exchange, it is helpful to refer to the notion, developed in cognitive linguistics, that is related in spirit to that of quotations as demonstrations, namely that of "fictive interaction" (Pascual, 2002, 2014). The central idea is that in fictive interaction, the frame of a face-to-face conversation is used in language to structure meanings that in fact involve no actual conversation. Examples include fictive speech acts as in *Call me old-fashioned, but…* (which are not 'actually' asking an interlocutor to call the speaker old-fashioned), and uses of apparent Direct Speech snippets at lower structural levels, like at the phrasal level, as in *a "yes we can" attitude*. What happens in such types of cases, then, is that something we know from real interactions—having a speaker, an addressee, a speech event in which they participate—is borrowed in order to represent mental construals such as emotional stances.

Our examples below, and also in section 3.5, illustrate several types of memes which rely on the structure of a dialogue to construct stance, while not always using fully profiled dialogic structures; also, such memes do not always represent viable conversational discourse. We consider such uses to belong to the category of 'simulated interaction', where memetic conversational-looking discourse structures do not just rely on evocation of interactions where specific stances are involved (which would apply broadly to fictive interaction examples). In the cases we consider below, the evocation of the dialogue form as such (rather than a specific discourse turn with accessible meaning such as "yes we can") is used to construct the representation of stance. Additionally, the construction of stance is supported via the very visual organization of the meme panels. In this section, we will consider three patterns which illustrate different strategies Meme Makers rely on.

We start by referring back to the Two Guys on a Bus meme. The contrasting stances there were allocated to two sides of the image macro, while there was no dialogic discourse added. The distinction between the stances was marked by different views outside the window of the bus, the facial expressions of the two men, and the labels identifying the target of positive/negative evaluation. For comparison, the Mad Men meme, illustrated in Figure 9, combines two stills from the drama series Mad Men, where a junior employee, with raised eyebrows, is shown saying "I feel bad for you" in the top image, and a more senior colleague in the bottom image, sharing an elevator ride after a pitch meeting, says "I do not think about you at all" while looking serious and frowning slightly. There is already text in the image macros: by convention, we read these lines of text ("I feel bad for you," "I do not think about you at all") as subtitles, and apportion them to the depicted speakers. On top of that, however, the meme adds labeling, thereby reusing the two-part image macro to apply the sequence of stance lead (pity/disappointment) and stance follow (rejection of the previous stance) to a range of topics—for instance pitting Europe against the US, or 'the 49 other states' against 'New Jersey'. In the example we include here (Figure 9), Android users are identified as expressing pity, whereas iPhone users are arrogantly dismissive.

Importantly, the top/bottom organization of the meme, and the not highly expressive faces and body postures weaken the assumed semblance of a conversational context (in the original scene, the characters are in fact standing next to each other in an elevator, both facing the door of the elevator, which explains the viewer-facing arrangement). In fact, the only indication of a discourse connection between the top and bottom panels is the pronoun *you*—the deictic pronoun identifying the addressee. The Meme Viewer has to create the interactive context needed solely on the basis of the content of the conversational lines and the use of deictic pronouns. It is worth noting, though, that deictic forms are a sufficient component to prompt the understanding of the entire meme in terms of simulated interaction.

A more complex, quite popular example is the Anakin and Padmé dialogue meme, using a four-cell grid to represent exchanges between two Star Wars characters (Figure 10). The basic grid structures a much more natural interactive pattern, with Anakin on the left-hand side, and Padmé on the right. The characters' eye-gaze also suggests interaction. Additionally, the changing stances are clearly signaled via facial expression: the shot of Anakin's face changes from the top left cell (where his line of text is "I'm going to

FIGURE 9
Mad Men meme. Reproduced from Imgur.com user RoyBattysDove, 8 September 2016.

FIGURE 10
Anakin and Padmé meme. Reproduced from Reddit.com user XipingVonHozzendorf, 22 May 2021.

change the world") to the bottom to become more close-up and with an intense expression on his face. There is no text in this bottom left cell. Meanwhile the image of the Padmé character changes drastically, from smiling and unconcerned in the top right cell, to troubled and sombre in the bottom cell, suggesting a very different intonation being used for the dialogue line, repeated in both of her cells, "For the better, right?." In this way, the left-to-right, top-to-bottom sequential ordering structures changing and evolving stances, in which the good intentions of the 'opening gambit' (in the top left cell) end up being called into question. Labeling and/or altering the text subsequently shifts the meaning of the original dialogue—which is more extensive in the actual film dialogue, and discusses forms of government (basically, democracy vs. dictatorship)—to the frame evoked by the textual and visual additions and alterations. In a simple case, as in Figure 10, the basic grid is not changed but labels are added in the four cells, here labeling Anakin as "Elon Musk" and Padmé as "The Internet" – presumably in a critique of Musk's influence on cryptocurrency markets by tweeting out conflicting information. Many varieties of the basic meme exist—for instance changing the dialogue lines, changing the pictures of people (with other pictures or even paintings, e.g., of Napoléon and Josephine; see Dancygier and Vandelanotte, 2025) or using visual labels (e.g., superimposing a flag to identify "Anakin" with a specific country). All these varieties typically respect some important essentials: the zoom-in for the Anakin character, and lack of a second dialogue line; the very stark change in stance facially expressed by the Padmé character, and the nature of the grid supporting the idea of conversational interaction. Textually, there is an added effect of the (near-)repetition of Padmé's dialogue line, the second iteration of it being uttered with strong tints of worry and disbelief, suggested by the embodied features. In this type of dialogic grid many elements combine to create stance—facial expressions represented, the nature of the default dialogic pattern, and the additional framing prompted by labeling. But, perhaps most importantly, by applying the dialogic pattern to a different frame (as with 'Elon Musk/the Internet'), the meme is used to express a stance or an opinion on the content of that frame, e.g., Musk's role in the changed perception of the Internet, and not to assume any spoken (fictive or otherwise) interaction between the entities profiled (Musk and the Internet). The attitudes and stances signaled via facial expressions and conversational quotes are applied to entities and situations in the frames, attributing stance to culturally rich concepts such as 'the Internet'. Discourse snippets and facial expressions in the grid are thus used to simulate attitudes in entities not represented in the memes as such.

A more recent example showing speech participants visually, and 'relabeling' their dialogue, is the David Beckham "Be Honest" meme. The meme presents a three-by-two grid showing stills from a documentary about former footballer David Beckham and his wife Victoria in which she (depicted on the left) talks about them both being "very working class," and he (shown on the right, listening in on his wife's interview from behind a door) insists repeatedly she should "be honest" until she relents and answers his question about what car her dad drove her to school in (a Rolls Royce, as it happens—not "very working class," then). On the "David" side of the grid, the bottom right image shows him, eyes closed and eyebrows raised, responding with "Thank you," in recognition of his wife finally properly conceding (as he sees it) his point. Figure 11 shows an example of the "Be Honest" meme, which maintains the text on the "David" side, but modifies the "Victoria" side to make it about something else entirely.

The form of the grid, with three tiers organized top-to-bottom, and gradually changing the stance represented (more insistence on the David side prompting more toning down on the Victoria side), additionally relies on the left/right allocation of conversational lines and facial expressions to the two participants in the conversation. It is enough to imagine the effect of any changes in the organization of the

grid (such as, putting Victoria on the right in the middle tier, or moving David's bottom tier image up top) to note that the conversation which viewers are asked to simulate is structured to represent the specific stance-shift happening throughout the conversation. The facial expression, the conversational turns and the nature of the grid jointly organize the simulated interaction involved.

The original exchange in the Beckham documentary is a fascinating piece of dialogue in its own right with the husband, listening in on his wife's interview from behind a door, intervening to force her to be more truthful. But it is perhaps even more fascinating to see how this was given a second lease of life as a meme in which people are forced to recognize that the things they say (e.g., "I wrote for 3 hours") often do not accurately reflect what is actually happening (e.g., hardly having written anything at all—a feeling all too recognizable to many writers). The meme is presumably not used for straight-faced, hard-hitting critique, as suggested by the smiling expression on David Beckham's face, or indeed the light tone of examples circulating on the Internet. Rather, it installs a mildly ironic, sympathetic distance (Tobin and Israel, 2012) from which to consider the human weakness in overstating one's credentials or achievements. Thus, proficient meme communicators who know what this particular artifact is used for, will have no difficulty identifying the emerging meaning and getting the meme's general point, regardless of the specific topics it is applied to: 'sometimes we all like to present ourselves as just that little bit better or more virtuous than we can actually truthfully claim to be.'

The examples we turn to in the next section add further dimensions of the concept of simulated interaction—including, for instance, the use of non-text lines (enclosed within asterisks) describing behavior, or indeed images, as turns in a dialogic exchange. These seem to us surprising and quite complex forms used in stance evocation, quite distinct from the use of existing conversational patterns (like in fictive interaction) if only because some of the forms centrally relied on do not actually appear in conversation at all.

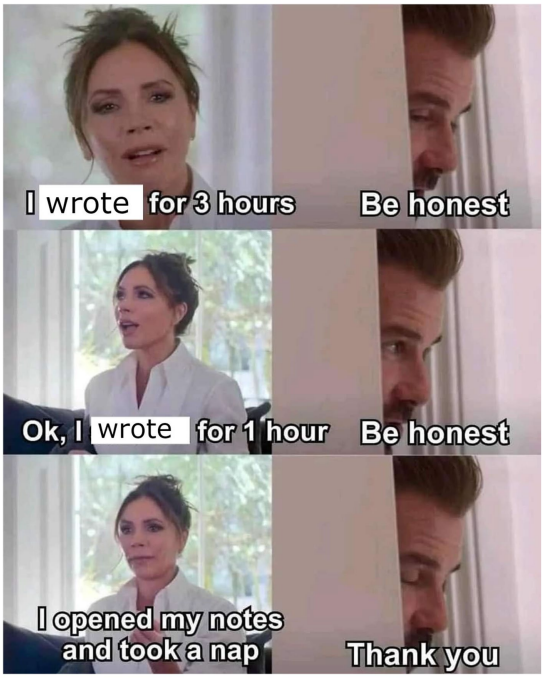


FIGURE 11
David Beckham "Be Honest" meme. Reproduced from X.com user AWrites116, 6 June 2024.

3.5 Simulated interaction formats not depicting speakers in memes and online discourse

We have seen the role that representations of embodied states and patterns of conversation play in simulated interaction. In this section we look more closely at the structure and functions of dialogic sequences in memes and other forms of online discourse which do not as such depict apparent 'interlocutors' in the way the Mad Men, Anakin and Padmé, and David Beckham "Be Honest" memes discussed in the previous section do. We will consider the role of pictures and emoji in different dialogue formats (using quotation marks in alternating lines, or using noun phrases to introduce different speakers). We will also discuss the use of the *Me* pronoun, referring to a Meme Character who represents the Meme Maker's experience. Finally, we will also highlight the absence of speech, and the staging of 'non-speakers,' in stance evocation.

We start with two examples that present turns in an apparent dialogue partly as discourse enclosed in quotation marks, and partly as pictures, which are profiled as turns or moves in conversational sequences. As such examples show, memetic discourse is not naturally interpretable in terms of familiar discourse categories such as 'narrative' or 'conversational'. In many cases, for example, fictive dialogic lines are used as representations of narrative events, and narrative sequence is represented by the sequence of lines in a faux dialogue used in the meme. As many of our examples show, such faux-dialogic lines can further be substituted with images, which have a higher humorous and stance-forming impact. Our contention is that the memetic re-construal of discourse types has two goals—(1) creating humor, while (2) simulating stance-loaded responses to the situations described.

Both dialogic and image-based construal can be seen in Figure 12 (@Stephenlough95 on X/Twitter, 24 December 2021), where the process presumably starts from the picture of a COVID self-test being

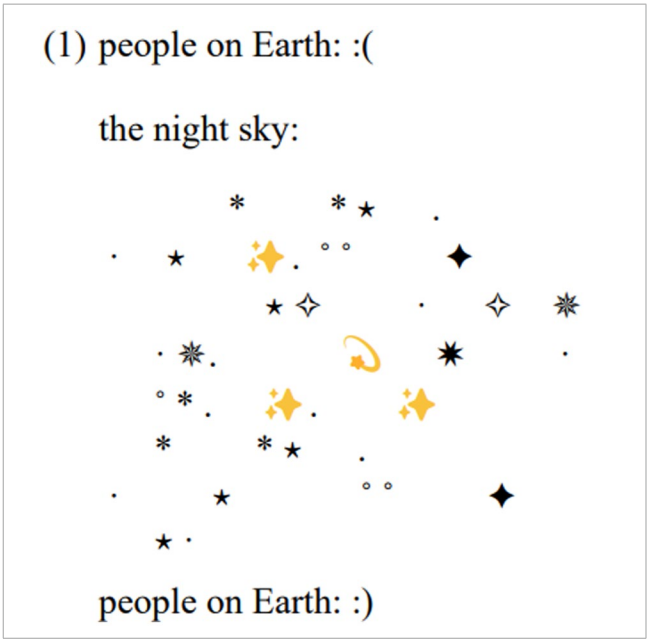
FIGURE 12
Faked Covid Test meme. Reproduced from X.com user Stephenlough95, 24 December 2021.

FIGURE 13
Bae Come Over meme. Reproduced from X.com user McJesse, 15 January 2018.

forged by adding in, in red pencil, the all-important second red line indicative of a positive test result. This picture was, for some time, popular on X/Twitter, and led several online communicators to imagine suitable fictive utterances, inviting someone to perform a duty or chore they would rather get out of, to which the picture provided the desired solution—getting out of something one does not want to do. The image in Figure 12 fictively responds to the fictive request to set the table by using an image as a token metonymically representing a verbal response such as "I cannot, I have COVID-19." Overall, the piece of simulated interaction communicates a stance about the task being avoided (dislike), but arguably also comments on the pandemic, making light of its new-found rituals, in this case regular self-testing.

Figure 13 (@McJesse on X/Twitter, 15 January 2018) shows a longer virtual exchange, with three lines of dialogue, which we are lexically prompted (by items such as "bae," for "babe," and talk of "parents" and "coming over") to understand as being between young lovers. The impossibility of coming over safely ("getting my brakes fixed") turns out to be swiftly ignored when the exceptional circumstance of absent parents is mentioned—the promise this holds appears to warrant extreme risk taking, but the picture shows the unlucky outcome in which "bae" crashed the car into a house along the way. The lines of fictive dialogue here are presented using quotation marks, without any accompanying reporting clauses to identify the speakers (whose generic identity we glean from context, as suggested above). This further underscores the fictive nature of the simulated interaction: no specific referents are involved at all, but a type of situation is pithily illustrated, including its disastrous outcome, exaggerated to the point of absurdity. The switch from a typical and realistic nature of the first three lines of the dialogue, to the highly evocative image representing a disastrous result creates humor through exaggeration. In a sense, such instances, similarly to *when*-memes, evoke a simile—a trope known for creating vivid and exaggerated imagery—so that the point of the joke is a suggestion that a young lover would prefer to risk certain death rather than miss the opportunity to be with someone they love.

Another strategy, very common on microblogging platforms such as X/Twitter, is that of adopting the type of dialogue format sometimes found in press interviews, in which different interlocutors are identified by means of noun phrases preceding a colon (*cf.* Vandelanotte, 2020, 2021; Dancygier and Vandelanotte, 2025). In Example (1), the rather broadly defined collective 'speaker' (people on Earth) changes their emotional stance completely after an intervention on the part of the starry 'night sky'—an intervention consisting in simply existing (and, we assume, in the communicator's view, being beautiful). The initial stance is represented by the basic sad face emoticon :(and the altered stance after consideration of the night sky by its counterpart, the basic happy face emoticon :). While the visual component here is really quite minimal (using emoji that are part of the available character set), we also have very extensive examples using only emoji. One such example (discussed in Dancygier and Vandelanotte, 2025) represents a Brexit negotiation between 'the UK' and 'the EU'—the two partners in the negotiation being represented as 'speakers' by means of their respective flag emoji preceding the 'speaker'-introducing colon—with the content of the exchanges being expressed solely using emoji, colons and arrows. An example such as this clearly shows how the general schematic form of a dialogue is used in quite original and unusual ways to evoke stance, and change in stance. Rather than take an existing form of possible 'real' dialogue and use it for another end, as in fictive interaction, here the apparent dialogue form itself is altered to allow things it does not actually allow in face-to-face conversations.



(@poetastrologers on X/Twitter, 13 April 2019)

The example in Figure 14 features use of the 'me' pronoun often found in memetic discourse to represent the Meme Maker's experience—for instance, of impatience. 'Me' is the memetic pronoun of choice, playing a role very much like a demonstrative marker—'what you see in the image represents my behavior/posture/facial expression/etc./ in the way that allows you (the Meme Viewer) to recognize the experiential or emotional stance I am describing'. The use of 'me' is a common feature of memetic discourse, as it allows the Meme Maker to represent their emotional stance in a somewhat impersonal and humorous way.

FIGURE 14
'Me' meme. Reproduced from Reddit.com user dalex001, 21 October 2017.

FIGURE 15
'Me/Person' Man on the Bench meme. Reproduced from Reddit.com user cej98, 19 July 2020.

Another very common pattern, the Me/Also Me meme, relies on clearly describing the Meme Maker's intentions or plans, only to show their failure to live up to them; one example we collected, for instance, shows as 'Me' line "I need to save money this month," and includes a picture of a woman with lots of shopping bags for the 'Also Me' part. In Figure 14, though, we see a faux-dialogic formula used to display types of behavior in response to recognizable everyday situations. In Figure 14, a parcel is scheduled for delivery (metonymically evoked by means of the fictive quote "your order has been shipped"), and the apparent dialogic response on the part of the 'me' faux-conversationalist is not anything verbal, but a depiction of the kind of expectant, hopeful attitude—as it happens, here embodied by the Meme Character of a dog sat on a chair and peering longingly out of the window.

For comparison, Figure 15 looks as if it is quite predictably structured as dialogue, with 'me', then 'person', then 'me' again. But in fact, each of these three lines is quite different, and none of them introduces speech, not even fictive speech. The first *me*-line describes what precedes the main event in focus—the Meme Maker's intention to leave a bench; the second line then introduces another Meme Character, "person," whose 'discourse' move is to appear on the scene, and take a seat on the bench. The final part of the meme uses the *me VERBing* pattern, describing the Meme Character (representing the Meme Maker) as waiting a bit before leaving, so as not to awkwardly suggest they are leaving *because* of the arrival of "person."[3] Here, a demonstration does follow the *me* phrase, even if it is not a verbal quote: a depiction of a man (actor Keanu Reeves) sitting on a bench sporting a somewhat blank expression. Even if, using standard conventions of dialogue in writing, the meme example here seems very complex and unusual, the artifact works because it builds on existing conventions of meme grammar—particular uses of pronouns (*me* typically representing the Meme Maker's views), roles such as Meme Maker and Meme Character, constructions such as *me VERBing* and its use with images. It also works because of the recognizability and relatability of the behaviors and stances described—we recognize the awkwardness of the situation, but we can also share with the Meme Maker a sense of the humor or even silliness of how we try to deal with this awkwardness.

We should also note here how this particular meme uses the faux-dialogic convention while using grammatical forms that would typically be considered unacceptable. First, 'Me:' is followed by a third person form *is*—which follows at least two memetic discourse conventions: using *me* to point to the Meme Character representing the Meme Maker's experience in the event described, and the third person *is* to follow the standard memetic usage which avoids talking about Meme Characters as if they were simply the same as the Meme Maker. In the next line, the text *sits next to me* is framed by asterisks, now a standard memetic convention to represent *action*. It is only in the final line, after the two characters in the story being told are established, that the Meme Maker can use the first-person pronoun to elaborate on their experience. Overall, this rather complex narrative meme uses faux dialogue in a rich way in order to simulate an emotional stance which explains superficially useless behavior (staying on when wanting to leave) in order not to hurt the feelings of a complete stranger—a situation and a stance that are relatable to the Meme Viewer.

The use of 'silent' lines, as with the night sky merely "being starry" in Example (1), or a person "sitting next to me" in Figure 15, is striking. Earlier, we also noted the use of silence as a 'conversational

3 For reasons of space, we do not elaborate our approach here in terms of mental spaces and blending (e.g., Fauconnier and Turner, 2002). For a treatment of many aspects of meaning-making in memes along these lines, we refer to Dancygier and Vandelanotte (2025).

turn' in the Anakin and Padmé memes (Figure 10)—the line (combined with Anakin's impassive stare) which makes Padmé realize that her assumption about Anakin's good intentions may be a mistake. A further related pattern was identified in Vandelanotte (2020), exemplified here in Example (2), in which a series of "empty" quotes, on the part of 'non-speakers' ("nobody," "absolutely no one," etc.) leads up to a culminating turn in which someone says something completely uninvited, and presented as irrelevant and unwanted. In the example, an 'iNfLuEnCeR'—the spelling reveals the online communicator's disdain for the mere category of people—starts a social media post as "A lot of you have asked about my skin care routine" (presumably continuing to advertise a particular brand of which they received freebies). The contrast between all the preceding lines of 'nothing being said or asked', and the influencer's fictive statement about "a lot of you" having asked about skin care, further underscores the negative emotional stance toward the way influencers operate. Interestingly, silent 'turns' in memes and other online discourse forms suggest online language users are acutely aware of the potential significance of silence in dialogue—a topic that is very actively explored in current approaches to conversation analysis (e.g., Hoey, 2020, 2021). Overall, (2) suggests an absence of speech, made expansive by dividing it over multiple lines of 'non-dialogue' on the part of 'non-speakers', followed by a stretch of unwanted, unasked-for speech—the influencer's intervention, presented as an annoying, self-centered and financially motivated form of engagement.

> (2) Nobody:
>
> Absolutely no one:
>
> Not a single soul on this Earth:
>
> Not even their mom:
>
> iNfLuEnCeR: "A lot of you have asked about my skin care routine..."
>
> (@cdcxpe on X/Twitter, 16 April 2019)

Note, again, in these examples how they deviate from existing conversational structure: the use of images, emoji, 'action' lines like **Sits next to me** presented where a dialogue line ought apparently to go, or a sequence of 'absent' dialogue lines from 'non-speakers' (*nobody, absolutely no one*, etc.) all do things which work in online visual-textual contexts, but not in the actual kinds of dialogic expressions the notion of fictive interaction is premised on. The more specific notion of simulated interaction is proposed to fill this gap.

4 Conclusion

We have tried to show, in this contribution, that despite their important differences, there is much that unites the 'embodied interaction' and 'polysemiotic' paradigms of multimodality research. Viewed from one perspective, the data types we have been presenting are far removed from face-to-face interaction (even 'mediated' online forms such as Zoom calls). At the same time, however, we see that even the snapshot depictions of embodied features—including facial expressions, gestures, (changes in) postures—however much reduced they may be compared to full face-to-face interaction, contribute much to the emergence of meaning, especially stance meanings, in internet memes. Likewise, the use of simulated dialogue formats in ways that are hard to imagine in face-to-face exchanges (presenting descriptions of behavior, for instance, or pictures, as 'turns' in the exchange) suggests interesting new ways of using the basic interactional frame of a dialogic exchange. Over the course of our exploration, we have suggested that online discourse is taking on specific features of embodied interaction, and developing strategies that constitute a new formula, which we have termed *simulated interaction*. Importantly, these communicative strategies specifically serve the needs of stance construction, the marking of contrasting stances, and also stance negotiation or shift. In other words, the wealth of stance phenomena online is developing expressive means which exploit the online reliance on visual means of expression and the need for brevity and clarity, while re-framing forms of interaction, both embodied and discourse-based. The new formulae of stance simulation call for an in-depth discussion, and this paper shows some potential directions worth exploring.

We have proposed the notion of simulated interaction as a more suitable concept, compared to fictive interaction, for the multimodal types of online discourse we have investigated. For one thing, some of our examples, such as the Distracted Boyfriend meme, concern not conversations, but purely embodied interactions (without a verbal component), which are sufficient prompts for Meme Viewers to construe stance exchanges. In addition, we have described a number of unique forms specific to online discourse which do not exist in 'actual' conversations and so cannot properly be analyzed as applying the existing 'conversation frame'. These include the use of images, emoji, descriptions of behavior (such as "Sits next to me"), or indeed of 'absent' speech by 'non-speakers', as apparent discourse moves. More broadly, we have seen various forms which blur the boundaries between narrative and conversation: examples such as the Bae Come Over meme or the 'Me/Person' Man on the Bench meme use long 'faux'-dialogic exchanges to in fact structure narrative sequences, with images representing either an emotionally loaded climactic event (Bae Come Over) or embodied features (Man on the Bench). Other examples, too, mix discourse snippets and narrative sequence (e.g., the It Will Be Fun, They Said meme, or the dialogic grid memes) in the pursuit of stance expression. We are thus proposing a concept which reflects the richness and complexity of online communication. It also relies in complex ways on two modes of multimodal discourse—multimodality in interaction (including gesture, body posture, facial expression, etc.) and image-text multimodality; specifically, images can be inserted in slots (such as dialogue entries) normally reserved for text, and play a broad array of stance evocation and narrative roles.

Overall, we have tried to show that much of the formal innovation observed in internet discourse is driven by a number of important discourse goals. The primary such goal is to represent experience, while not making the artifacts uniquely connected to the internet user's life and beliefs—hence, for instance, the appeal to generic pronouns or noun phrases such as "person," or the drive to connect experiences between unrelated frames and situations (e.g., a distracted boyfriend and a failure to follow a diet). Much of what is being communicated is opinions, reflections, humorous commentary, irony,

and sometimes sheer amazement at things we say and do to maintain the social fabric, while being aware of our own flaws and quirks. Humor itself is also a goal, as is brevity or 'linguistic economy'—if a couple of faux dialogic lines will do, why tell a long story? In this pursuit of pithiness, the use of images is especially effective, and as we have seen, images are used liberally to represent experience, and they can be naturally embedded in what looks like ordinary dialogue. These strategies of stance expression form the grounding of what we have termed 'simulated interaction'. Further research will, we hope, throw more light on the ways in which online communicators respond to, but also co-construct, instances of stance expression through simulated interaction (see, e.g., Dancygier and Vandelanotte, 2025, Chapter 10).

Data availability statement

Publicly available datasets were analyzed in this study. The data analyzed in this study is part of an ongoing collection of examples, collected by both authors from various public online sources over the years, for purposes of qualitative (not quantitative) analysis. The same or similar examples can be found on meme collecting sites, social media platforms or using online search engines. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving human data in accordance with the local legislation and institutional requirements. Written informed consent was not required, for either participation in the study or for the publication of potentially/indirectly identifying information, in accordance with the local legislation and institutional requirements. The social media data was accessed and analyzed in accordance with the platform's terms of use and all relevant institutional/national regulations.

Author contributions

BD: Writing – original draft, Writing – review & editing, Conceptualization. LV: Writing – original draft, Writing – review & editing, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding allowing open access publication was provided by The University of British Columbia, Canada.

Acknowledgments

Our thanks to the editors of the Research Topic on Stance-Taking in Embodied and Virtual Interaction for inviting us to contribute, and to reviewers for valuable feedback.

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

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P. (2023). Multimodal stance-taking in interaction – a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Attardo, S. (2023). Humor 2.0: How the internet changed humor. London: Anthem Press.

Barrett, L. F. (2017). How emotions are made: The secret life of the brain. Boston & New York: Houghton Mifflin Harcourt.

Bergen, B. (2012). Louder than words: the new science of how the mind makes meaning. New York: Basic Books.

Bülow, L., Merten, M.-L., and Johann, M. (2018). Internet-Memes als Zugang zu multimodalen Konstruktionen. *Zeitschrift Angewandte Linguistik* 2018, 1–32. doi: 10.1515/zfal-2018-0015

Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026

Clark, H. H., and Gerrig, R. J. (1990). Quotations as demonstrations. *Language* 66, 764–805.

Couper-Kuhlen, E., and Selting, M. (2017). Stance and footing. Supplementary online chapter "C" Of interactional linguistics: Studying language in social interaction. Cambridge: Cambridge University Press.

Dancygier, B. (1998). Conditionals and prediction: Time, knowledge and causation in conditional constructions. Cambridge: Cambridge University Press.

Dancygier, B. (2012). "Negation, stance verbs, and intersubjectivity" in Viewpoint in language: A multimodal perspective. eds. B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 69–93.

Dancygier, B. (2017). Viewpoint phenomena in constructions and discourse. *Glossa* 2:37. doi: 10.5334/gjgl.253

Dancygier, B. (2021). Fictive deixis, direct discourse, and viewpoint networks. *Front. Commun.* 6:624334. doi: 10.3389/fcomm.2021.624334

Dancygier, B., and Vandelanotte, L. (2016). "Discourse viewpoint as network" in Viewpoint and the fabric of meaning: Form and use of viewpoint tools across language and modalities. eds. B. Dancygier, L. Wei-lun and A. Verhagen (Berlin & Boston: De Gruyter Mouton), 13–40.

Dancygier, B., and Vandelanotte, L. (2017a). Viewpoint phenomena in multimodal communication. *Cogn. Linguist.* 28, 371–380. doi: 10.1515/cog-2017-0075

Dancygier, B., and Vandelanotte, L. (2017b). Internet memes as multimodal constructions. *Cognit. Linguist.* 28, 565–598. doi: 10.1515/cog-2017-0074

Dancygier, B., and Vandelanotte, L. (2025). The language of memes: patterns of meaning across image and text. Cambridge: Cambridge University Press.

Dawkins, R. (1976). The selfish gene. Oxford: Oxford University Press.

Denisova, A. (2019). Internet memes and society: Social, cultural, and political contexts. New York: Routledge.

Droz-dit-Busset, O. (2022). "So-called influencers": Stancetaking and (de)legitimation in mediatized discourse about social media influencers. *Discourse Context Media* 49:100629. doi: 10.1016/j.dcm.2022.100629

Du Bois, J. W. (2007). Stancetaking in discourse: Subjectivity, evaluation, interaction, 139–182. Amsterdam and Philadelphia: John Benjamins.

Fauconnier, G., and Turner, M. (2002). The way we think: Conceptual blending and the Mind's hidden complexities. New York: Basic Books.

Feyaerts, K., Brône, G., and Oben, B. (2017a). "Multimodality in interaction" in The Cambridge handbook of cognitive linguistics. ed. B. Dancygier (Cambridge: Cambridge University Press), 135–156.

Feyaerts, K., Oben, B., Lackner, H. K., and Papousek, I. (2017b). Alignment and empathy as viewpoint phenomena: the case of amplifiers and comical hypotheticals. *Cognit. Linguist.* 28, 485–509. doi: 10.1515/cog-2016-0109

Goodwin, M. H., Cekaite, A., and Goodwin, C. (2012). "Emotion as stance" in Emotion in interaction. eds. A. Peräkylä and M.-L. Sorjonen (Oxford: Oxford University Press), 16–41.

Hoey, E. M. (2020). When conversation lapses: The public accountability of silent Copresence. Oxford: Oxford University Press.

Hoey, E. (2021). "Sacks, silence, and self-(de)selection" in On sacks: methodology, materials, and inspirations. eds. R. J. Smith, R. Fitzgerald and W. Housley (London: Routledge), 130–142.

Jenkins, E. S. (2014). The modes of visual rhetoric: circulating memes as expressions. *Q. J. Speech* 100, 442–466. doi: 10.1080/00335630.2014.989258

Kang, J.-i., Hanbeom Jung, A., Kwon, Y., and Kwon, I. (2023). Meaning extensions of internet memes: a case study of the 'if 2020 was a(n) X' meme. *Rev. Cognit. Linguist.* 21, 178–209. doi: 10.1075/rcl.00131.kan

Lou, A. (2017). Multimodal simile: the "when" meme in social media discourse. *English Text Construct.* 10, 106–131. doi: 10.1075/etc.10.1.06lou

Lou, A. (2021). Multimodal similes: Toward a cognitive understanding of Similative meaning. Vancouver: PhD dissertation, The University of British Columbia.

Milner, R. M. (2016). The world made meme: Public conversations and participatory media. Cambridge, Mass: The MIT Press.

Ochs, E. (2006). "Linguistic resources for socializing humanity" in Rethinking linguistic relativity. eds. J. J. Gumperz and S. C. Levinson (Cambridge: Cambridge University Press), 407–437.

Pascual, E. (2002). Imaginary trialogues: Conceptual blending and fictive interaction in criminal courts. Utrecht: LOT.

Pascual, E. (2014). Fictive interaction: The conversation frame in thought, language, and discourse. Amsterdam & Philadelphia: John Benjamins.

Pascual, E., and Sandler, S. (Eds.) (2016). The conversation frame: Forms and functions of fictive interaction. John Benjamins: Amsterdam & Philadelphia.

Paz, M. A., Mayagoitia-Soria, A., and González-Aguilar, J.-M. (2021). From polarization to hate: portrait of the Spanish political meme. *Social Media Society* 7. doi: 10.1177/20563051211062920

Piata, A. (2020). Stylistic humor across modalities: the case of classical art memes. *Internet Pragm.* 3, 174–201. doi: 10.1075/ip.00031.pia

Ross, A. S., and Rivers, D. J. (2017). "Internet memes as polyvocal political participation" in The presidency and social media: Discourse, disruption and digital democracy in the 2016 presidential election. eds. D. Schill and J. A. Hendricks (New York: Routledge), 285–308.

Ruppenhofer, J., and Michaelis, L. A. (2010). A constructional account of genre-based argument omissions. *Construct. Frames* 2, 158–184. doi: 10.1075/cf.2.2.02rup

Shifman, L. (2014). Memes in Digital Culture. Cambridge, MA: The MIT Press.

Sweetser, E. (2012). "Introduction: viewpoint and perspective in language and gesture, from the ground down" in Viewpoint in language: A multimodal perspective. eds. B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 1–22.

Talmy, L. (1988). Force dynamics in language and cognition. *Cogn. Sci.* 12, 49–100. doi: 10.1207/s15516709cog1201_2

Tobin, V., and Israel, M. (2012). "Irony as a viewpoint phenomenon" in Viewpoint in language: A multimodal perspective. eds. B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 25–46.

Vandelanotte, L. (2019). Changing perspectives: something old, something new. *Pragmatics* 29, 170–197. doi: 10.1075/prag.18046.van

Vandelanotte, L. (2020). (Non-)quoting and subjectivity in online discourse. *E-rea* 17.2. doi: 10.4000/erea.9782

Vandelanotte, L. (2021). Creative constructs, constructions, and frames in internet discourse. *Construct. Frames* 13, 160–191. doi: 10.1075/cf.00050.van

Vandelanotte, L., and Dancygier, B. (2017). Multimodal artefacts and the texture of viewpoint. *J. Pragmatics* 122, 1–9. doi: 10.1016/j.pragma.2017.10.011

Veale, T. (2012). Exploding the creativity myth: The computational foundations of linguistic creativity. London: Bloomsbury.

Walker, Aidan. (2023). How to read a meme: Context, agency and knowledge in distracted boyfriend. New York: MA thesis, Columbia University.

Wiggins, B. E. (2019). The Discurse power of memes in digital culture: Ideology, semiotics, and intertextuality. London: Routledge.

Wiggins, B. E., and Bower, G. B. (2015). Memes as genre: a structurational analysis of the memescape. *New Media Soc.* 17, 1886–1906. doi: 10.1177/1461444814535194

Wikström, P. (2019). Acting out on twitter: affordances for animating reported speech in written computer-mediated communication. *Text Talk* 39, 121–145. doi: 10.1515/text-2018-2021

Xie, C. (2022). The pragmatics of internet memes. Amsterdam: John Benjamins.

Yus, F. (2018). Identity-related issues in meme communication. *Internet Pragmatics* 1, 113–133. doi: 10.1075/ip.00006.yus

Zappavigna, M., and Logi, L. (2024). Emoji and social media paralanguage. Cambridge: Cambridge University Press.

Zenner, E., and Geeraerts, D. (2018). "One does not simply process memes: image macros as multimodal constructions" in Cultures and traditions of wordplay and wordplay research. eds. E. Winter-Froemel and V. Thaler (Berlin and Boston: De Gruyter Mouton), 167–193.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 12 March 2025
DOI 10.3389/fpsyg.2024.1436677

Check for updates

OPEN ACCESS

EDITED BY
Geert Brône,
KU Leuven, Belgium

REVIEWED BY
Maria Grazia Rossi,
NOVA University of Lisbon, Portugal
Isabella Poggi,
Roma Tre University, Italy

*CORRESPONDENCE
Aija Logren
✉ aija.logren@uef.fi

†These authors have contributed equally to this work and share first authorship

RECEIVED 22 May 2024
ACCEPTED 02 September 2024
PUBLISHED 12 March 2025

CITATION
Logren A, Ilomäki S and Ruusuvuori J (2025) Management of divergent stances as a resource to maintain progressivity and social relationships.
Front. Psychol. 15:1436677.
doi: 10.3389/fpsyg.2024.1436677

COPYRIGHT
© 2025 Logren, Ilomäki and Ruusuvuori. 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.

Management of divergent stances as a resource to maintain progressivity and social relationships

Aija Logren[1*†], Sakari Ilomäki[2†] and Johanna Ruusuvuori[2†]

[1]Department of Social Sciences, Faculty of Social Sciences and Business Studies, University of Eastern Finland, Kuopio, Finland, [2]Faculty of Social Sciences, Tampere University, Tampere, Finland

Previous studies have shown the intersubjective and negotiable nature of stance: interlocutors orient to alignment and adjust their stances to achieve closer alignment. In this article, we study the interplay of three axes of stance—epistemic, deontic and affective stance—and the role their management may have in socially relevant tasks. We describe how the three axes can be simultaneously relevant, taken into account, and dynamically shifted by the participants in a specific sequence of action. The three axes are not always equally aligned or disaligned, but instead divergent: some are aligned at the same time when others are disaligned. Through a case study with two data excerpts, we show how the divergence is an interlocutors' resource to overcome the disalignment of some of the stances, and to eventually achieve sufficient alignment in order to proceed their activity. Our data are drawn from the institutional context of neurological consultations. We examine the interactants' stance over longer episodes of talk to illustrate their momentary, multimodal interactional work to display and adjust their stances. The interactants deploy different modalities to address the divergent stances, and further, the multimodal and multifaceted nature of turns enable them to orient to several axes of stance at the same time. Instead of merely taking a stance, the interlocutors *manage their stances*—both in terms of adjusting the alignment and the balance of the different axes—and thus maintain the social relationship between themselves and the progressivity of the ongoing task.

KEYWORDS

management of stance, stance-taking, epistemics, deontics, affect, divergence, asymmetry

Introduction

In any human interaction, people orient to maintaining their shared understanding about the world (Clark, 1996). This involves paying attention to three omnipresent axes of meaning making that are relevant in creating and maintaining this 'common ground': knowledge, power and emotion, i.e., the epistemic, deontic and affective orders of interaction (Stevanovic and Peräkylä, 2014). These three axes become more or less salient in the ways in which we project and recognize different actions in interacting with each other, depending on the type of action that is ongoing (Levinson, 2012). In constructing an action the interactants bring forward their *stance* toward the object or topic in focus. In this article, we describe, firstly, the dynamics of simultaneous, yet potentially divergent epistemic, deontic and affective stances and secondly, introduce a theoretical concept, *management of stances*, to illuminate the momentary interactional work interactants do to display and adjust their stances. Our data are drawn from the institutional context of neurological consultations where participants' differing roles in terms of the institutional task help to highlight the dynamics of this management work.

Intersubjectivity of stances has been described with a theoretical framework 'Stance Triangle' by Du Bois (2007), which illustrates how stance is not merely individual and subjective evaluation of objects, but an intersubjective process. When taking a stance, a person (P1, see Figure 1) evaluates an object (O), and in so doing, positions themself in relation to that object. In other words, by taking a stance, the person displays that they are, for example, able, entitled or obligated to do so (Heritage, 2012; Stevanovic, 2018; Stevanovic and Peräkylä, 2012; Stivers et al., 2011). Nonetheless, because the stances are not displayed in isolation, the interactants' responses to P1's stance-taking contribute to it as well. Interactants' (P2) responses, which also display evaluation of and positioning to the object, i.e., a stance of their own, are either in alignment to the P1's stance, or disalign with it (Du Bois, 2007). Alignment and disalignment are consequential for the relationship of the interactants, and thus they may finely adjust their subjective stance displays in their following turns and embodied actions to achieve alignment (see Kaukomaa et al., 2013; Ruusuvuori and Peräkylä, 2009).

We introduce a notion that all three stances—epistemic, deontic and affective stance—can be simultaneously relevant and taken into account by the participants in a specific sequence of action. Furthermore, they may not always be equally aligned or disaligned. Instead, there may be a divergence of the stances, for example the epistemic stance could be disaligned simultaneously when the deontic and affective stances could be aligned (Figure 2).

Because of the possible divergence of stances and their alignment, the fine-grained management is crucial: stances are not static, but they, and perhaps their balance, are finely adjusted and shifted in interaction (Logren et al., 2019, 2020; Ruusuvuori and Peräkylä, 2009). Hence, we argue that the theoretical concept "stance-taking" does not capture the complex and dynamic interactional work participants do in displaying and adjusting their stances and the alignment of the stances. Building on the notions by Küttner (2019), Tai et al. (2022), and Sakita (2013, 2017), who have described how interactants employ specific types of meta-talk to index differences in stances and to negotiate and repair their differences such as disagreement, we suggest that examining how stances are managed over longer episodes of talk better illuminates the intersubjective nature of stance-taking and its social relevance. In this article we illustrate how management of stances is a participants' resource to maintain progressivity and social relationships.

We present a case study of the divergence of the three stances, and their momentary management through verbal and embodied means in an institutional context, namely clinical consultations. More specifically, drawing from two extracts from neurology outpatient clinics where patients voluntarily bring forward their own understanding about the causes of their symptoms or some suggested treatments, we analyze how the patients display epistemic, deontic and affective stances, how the doctors orient to these different stances in their responses, and how both participants work to adjust the alignment and divergence of their stances.

Conversation analytic perspective on epistemic, deontic and affective stance

Within our theoretical perspective to the three axes of stance and their alignment and divergence, we adopt the framework of Conversation Analysis (CA) to entangle the momentary sequential work interactants do to display and adjust their stances, participatory roles and preference. This enables us to describe how alignment and divergence of stances is managed turn by turn, and the consequences this may have to the progressivity of the interaction and to the relationships of the interactants. However, in CA alignment has been used also to describe *structural alignment*, that is, how the following turns of talk continue the course of action taken in previous turn (see Stivers et al., 2011). The *alignment of stances* in the sense that has been described in the theory of Stance Triangle (Du Bois, 2007), and thus is the backbone of our study, has been depicted in previous CA studies with concepts such as agreement, acceptance and affiliation. In the following, we review the perspectives to the three axes of stance previous CA studies have to offer, and reflect their relationship to our theoretical framework of the three parallel axes of stance: epistemic, deontic and affective stance—that is, knowledge, power and emotion.

In CA, the knowledge dimension of interaction is referred to as *epistemics* (e.g., Heritage and Raymond, 2005). Management of knowledge in interaction can be differentiated to epistemic stance, which refers to interactional practices through which participants display their evaluation and positioning toward the domain of knowledge, and to epistemic status, i.e., how knowledgeable one participant is entitled to be about some domain of knowledge (Heritage,

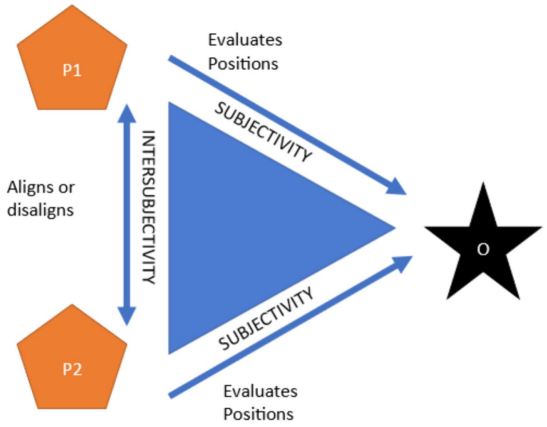


FIGURE 1
The Stance Triangle (Du Bois, 2007).

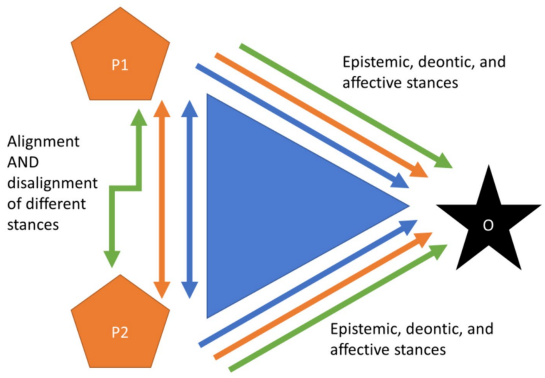


FIGURE 2
Divergence of stances: simultaneous alignment and disalignment of the different stances within the interlocutors.

2012). Alignment with the previous speaker's epistemic stance, that is, agreement, is achieved with specific components such as repetition, upgraded evaluations and explicit confirmations, and is often produced with minimal gaps, whereas disalignment, i.e., disagreements are often prefaced, delayed and mitigated (Du Bois, 2007; Heritage and Raymond, 2005, see also Pomerantz, 1984). Furthermore, because of a general preference for agreement in interaction (Sacks, 1987), interlocutors face a practical problem to solve—in Heritage and Raymond's (2005, p. 15) words: "whose view is the more significant or more authoritative"—whenever they disagree with the previous speaker, i.e., when their epistemic stances disalign. Interlocutors take into account the relationship between their epistemic stance and status, which can be observed in practices such as downgrading and upgrading one's access to the object of knowledge. For example, questions bring forward some pre-assumptions the questioner has (Heritage, 2009) and turn design and choice of words can be used to convey that some information is known and certain to the recipient—or the opposite (see for example Heritage, 1984, the use of particle "Oh" in English).

The power to make things happen, either in the immediate interactional situation or in the future, is approached in CA through the concept of *deontics* (Ekström and Stevanovic, 2023; Stevanovic, 2018; Stevanovic and Peräkylä, 2012). Deontics refer not only whether an individual gets others to act as they wish but more importantly, how these others consider the individual's right to imply some future actions in the first place. In many ways, same kind of general principles apply to the management of deontic dimension in interaction as in the case of epistemics. Deontics can be conceptualized through deontic status (i.e., the ability, entitlement or responsibility to make something happen) and deontic stance (i.e., how that is actualized in interaction). Should the recipient affirm the portrayal of the speaker's deontic right, deontic congruence is formed and vice versa, deontic incongruence exists when the deontic stance is challenged (Stevanovic and Peräkylä, 2012). The production of this (in)congruence is actualized in the alignment or disalignment of interlocutors' deontic stances, i.e., in acceptance or rejection of previous speaker's stance. Thus, deontic authority is not a static phenomenon but something that is actualized in interaction: deontic authority varies from domain to domain (that is, a person might have authority about some decisions while not other), and while some status bases for authority may exist, the authority can be resisted in interaction (Stevanovic and Peräkylä, 2012). Disalignment of deontic stances again induces a practical problem for the interlocutors: how to continue, if both parties hold on to their stance. To solve—and to preempt—this dilemma, participants upgrade and downgrade their deontic authority, for example through fitting their demands for action to their entitlement to command and others' readiness to comply (Antaki and Kent, 2012; Curl and Drew, 2008).

Within the framework of CA, the emotional dimension of interaction becomes salient when the participants orient to some feature of the conversation as affective (Local and Walker, 2008). The affective dimension differs from the epistemic and deontic ones in that it can be (and often is) displayed also, or only non-verbally, with various multimodal means and gestalts. Another difference is that affective status cannot be contested by the other interlocutors—Sacks (1992, Vol II, 242–243) talks about entitlement to experience, meaning that witnessing an event and enjoying or suffering from the experience caused by it, makes a person more entitled to know and tell about it than someone who has only second-hand information about the incident in question. Thus, notwithstanding potential differences in epistemic or deontic status, affective status remains with the person who 'owns the experience' (term coined by Peräkylä and Silverman, 1991). Affective stance, then, refers to the affective evaluations that are displayed in interaction and that display the interlocutors' emotional orientation (Goodwin et al., 2012). Although emotion may not be made lexically explicit in conversation, it is omnipresent in the sense that any act in interaction may be interpreted as displaying some affective stance. For example, eye-rolls are interpreted as displaying dissent toward the immediately preceding action (Clift, 2021), and turn-initial frowns foreshadow complications within the upcoming story-telling (Kaukomaa et al., 2013). Even the lack of showing affect when it would be relevant is also a display of an affective stance: for example, failing to validate other interlocutor's troubles-telling may signal lack of empathy and lead to pursuing validation (Ruusuvuori, 2005). Alignment or disalignment (i.e., affiliation or disaffiliation) of affective stances may be less crucial to the progressivity of the ongoing interaction as affective stance displays are often non-verbal and do not necessarily follow the sequential order of verbal interaction (cf. Mondada, 2018). Thus, while affective stances may be reciprocated (as in validating a troubles-telling), they may also be left unnoticed or untopicalized (as in responding to troubles-telling with advice (see Jefferson and Lee, 1992). Nevertheless, disalignment may have consequences for the social relationship of the interlocutors (Du Bois and Kärkkäinen, 2012).

Analytic interest has typically focused on one axis at a time (which is reflected also in the differing terminology described above), but our initial empirical observation has been that interlocutors can orient to all three at the same time, and moreover, from different perspectives at the same time. Therefore, it is necessary to examine, firstly, *how* they accomplish this complex, threefold orientation: what kind of features of turns enable the three stances to intertwine, and how interlocutors deploy different modalities to address the three axes and their divergence. This is also the reason we have chosen the analytical concept *alignment*, which depicts the parallel and intertwined relationship of the three stances. Secondly, *why* this threefold orientation seems to be relevant for the participants? One reason for this may lie in the context of the interaction. The roles and relationships of the participants may be particularly salient in specific types of contexts, e.g., in formal, hierarchical institutional contexts such as clinical encounters. Thus, the epistemic, deontic and affective statuses and managing the three stances becomes specifically relevant for the interlocutors. Furthermore, the tasks of the encounter the participants pursue may increase the importance of the progressivity of the encounter.

Clinical encounters as a context for managing stances

Clinical encounters form a specific kind of institutional and interactional context, which shapes the production of all three, epistemic, deontic and affective aspects in interaction: We analyze cases from neurology consultations, as with the institutional task of finding proper medical treatment for the patient, and the pre-assigned expertise of the professionals, there is a clear asymmetry between the rights and obligations of the participants. This means that the participants' epistemic and deontic statuses remain stable as compared to everyday conversation. This also means that patients may have to do additional interactional work to make their own concerns heard, which then makes negotiation over disaligned stances relevant (for overviews on institutional interaction, see Drew and Heritage, 1992; Heritage and Clayman, 2010).

Regarding the epistemics, there is a basic orientation to two domains of knowledge and their primary epistemic authorities: the medical knowledge, for which the professional is the authority, and the patient's lifeworld knowledge, which includes for example bodily sensations, earlier experiences with treatments, and preferences about different solutions. While these domains exist, their boundaries are not strict in a sense that for example the patient could not express their understanding about the causes of symptoms. Instead (Curl and Drew, 2008) when patients produce their ideas, they balance between commitment to the idea and demanding an uptake from the professional (Gill, 1998; Gill and Maynard, 2006). Thus, both patients and professionals generally produce turns that are in congruence with their epistemic status: patients by lowering their epistemic stance (Gill and Maynard, 2006) and professionals by maintaining theirs (e.g., Peräkylä, 1998, see Stivers et al., 2011 on epistemic (in)congruence).

The clinical setting provides structure for the management of deontics, regarding both progression within the consultation (e.g., Robinson and Stivers, 2001) and decision-making (Land et al., 2017). While only the professional has the right to prescribe treatments, the patient always has the option not to comply, and for example refrain from taking the prescribed medicine. Thus, some kind of display of commitment is crucial when participants make decisions in clinical interaction (Stivers, 2006). Due to the institutional context where doctors have the deontic rights to prescribe medication, patients' ways to resist the decision are usually subtle. Patients may, for instance, ask questions on possible other medication or treatment, or refrain from committing to the treatment suggestion by mere nods (Stivers, 2002). They may also resort to their undeniable epistemic rights to know more about their current symptoms.

In clinical interaction, as well as undoubtedly in many other institutional contexts, affective stance is subordinate to epistemic and deontic stances which are focal in regards fulfilling the institutional purpose of the encounter. For example, while the patients may produce affective turns which in everyday conversation would make relevant an equally affective response, in clinical context it is often relevant to maintain the orientation to the task (Jefferson and Lee, 1992). However, the affective axis may still be oriented to as relevant, especially by the patients who may at times pursue validation to their troublesome experiences treating the professional's task-oriented response as insufficient (Ruusuvuori, 2005, 2007). Furthermore, while epistemic and deontic authority may become a subject of contest, affective status cannot be contested in the same way by the other interlocutors—instead, the patient is treated as entitled to their own experience (Peräkylä and Silverman, 1991). Thus, while patients' deontic status in deciding which treatment is appropriate is low as compared to the doctor, they may resort to their affective experiences in resisting the doctor's decisions [as in resisting the doctor's proposal to go to hospital by claiming to be afraid of hospitals, (see Ruusuvuori, 2005)].

As this last notion demonstrates, epistemics, deontics and affect are not separate entities, and it has been suggested that especially the first two intertwine in clinical interaction (Heritage, 2013; Stevanovic, 2018; Stevanovic and Peräkylä, 2012). For example, patients can treat medical professionals' assertions about available treatments either as suggestions or as mere informings (Toerien, 2018) and when acceptance of a suggestion is expected, patients' questions can be treated as a way of doing rejection (Stivers, 2005). Thus, whether and how epistemics, deontics or affect are relevant for the participants must be determined *in situ*. However, while the relationship between epistemics and deontics has gained interest, especially from the viewpoint on how epistemic authority can be used to build deontic authority (see, e.g., Peräkylä, 1998) and how knowledge configures in decision-making (see, e.g., Toerien, 2018), the management of all these three dimensions simultaneously has received less attention (see however, Koskinen and Stevanovic, 2022; Logren, 2019; Logren et al., 2019; Petraki and Clark, 2016 on how epistemic and affective stance intertwine). In this article, we examine how the three stances can have parallel relevance for the interactants, and how some of the different stances may be simultaneously aligned and some disaligned, that is, divergent.

Materials and methods

The data are drawn from a corpus of 86 neurology outpatient clinic consultations, collected in 2021–2024 (data collection ongoing) in two hospitals. The participants are patients who are visiting neurology clinics for diagnosis or treatment adjustments and their doctors and nurses. The data are in Finnish. The consultations were video-recorded and transcribed using Jeffersonian transcript symbols (Jefferson, 2004) that were augmented to depict multimodal actions (Mondada, 2001, see Supplementary material for transcription symbols and original Finnish transcripts).

During unmotivated looking at the data, we recognized those patients' tuns, in which they put forward their understanding and ideas about the causes of some symptoms or treatments to be interesting interactional situations regarding how the shared understanding and the alignment of epistemic stances is achieved step by step. In parallel with analyses concentrating on epistemics, we recognized how in some cases also deontic and/or affective stances were made relevant. We then concentrated on the cases where multiple stances are made relevant. Finally, we selected two contrasting cases for the analysis of this article to showcase the complexity of managing three stances simultaneously: one depicting a simple case focusing only on epistemics, and one with all three stances being made relevant (see Schegloff, 1987, on conversation analytic approach to single-case analysis).

The study is part of "Reliable Knowledge for Health Care: Process and Practice of Shared Decision Making" research project, funded by the Strategic Research Council (SRC) established within the Academy of Finland (project numbers 31213358415 and 31213584181). Before the data collection, the research project obtained ethical approval from the local ethics committee for medical research in Pirkanmaa region, Finland (reference code R21057), organizational permissions from the participating hospitals and signed informed consent from the participants. The data extracts are pseudonymized.[1]

Results

Our results are twofold. On a more theoretical level, we demonstrate how epistemic, deontic and affective stances can co-exist while being divergent—that is, participants' stances can align on some of the three while simultaneously disalign on other(s). This

1 The authors want to thank Pinja Majava for producing the line drawings used in this article.

phenomenon is potentially subordinate to the institutional task at hand. On a more empirical level, we show how different modalities can be used to manage the different stances and their alignment. Furthermore, by achieving alignment in one axis it is possible to mitigate disalignment in the other two. Thus, the management of divergent stances appears to manage the relationship and the progress of the situation.

Our first extract depicts a simple case where the participants manage primarily the epistemic axis. The extract is 22 min into the consultation: based on the history-taking the participants have decided that the doctor will consult a colleague to discuss the case before setting a diagnosis. Right before we join the action the doctor has initiated the transition to physical examination (Figure 3).

While the patient initially passes the opportunity to ask a question (lines 1–3), after the doctor's continuer (line 5) he starts to produce his idea that the vaccination for the SARS-CoV-2 virus could be the reason for his symptoms. The patient frames the idea as something that has occurred to him spontaneously without his own effort (cf. Halkowski, 2006; Ruusuvuori, 2000, line 6), and provides the grounds for the idea by evaluating the potentially close timing of the vaccination and his symptoms (lines 7–10), referring to his symptoms vaguely as "this hustle." As the doctor merely receives the idea as information with *yes* (line 11) the patient continues by explicating the idea (line 12). The patient brings his idea forward in the form of a question, making the doctor's uptake relevant. In addition to framing the idea as passively occurring, the question format also enables the patient to bring this idea forward without strong public commitment to it (compared to, for example, stating that he thinks the covid vaccination has caused the increase of the symptoms, cf. Gill, 1998).

Through these practices the patient produces his epistemic status as low and his stance as uncertain. Furthermore, he does not pose the idea straightforwardly but only after passing the first opportunity to

```
   d: >>stands up, moves towards the patient->
   p: >>gaze towards D->
 1 L: would you at this point have questions in +%your  +mind.
   d:                                             %turns towards the Pat.->
   p:                                          ->+,,,,,,,+gaze forward->
 2      (0.4)%(0.2)
   d:     ->%standing facing the patient, back towards the camera->*
 3 P: .hhh >well I quess I don't have< likeh,
 4     (0.3)
 5 D: ↑yes.
 6 P: in that% sense of course it has- .hh >it has come to mind
   d:         %moves sideways, blocking the camera
 7 P: that< (0.8) like like (0.9) in August I got that last (.)
 8 P: third (0.3) Covid  vaccination  so li[ke   ] (0.3) like maybe
 9 D:                                    [/mm-m.]
10 P: now after that there has been then a bit more of this hustle then
like.
11 L: yes.
12 P: .so .like (.) could it possibly from that then (0.2)
13 P: [ come    (--)          ]
14 D: [ .hh I- I don't believe] that they link
15 D: [to that covid vaccination.     ]
16 P: [>maybe they don't< link to that] yes I don't believe that either
17 P: [like]
18 D: [mm. ]
19 P: .yes
20     (0.4)
21 D: so: I: wouldn´t (.) I wouldn't consider it very pro[bable]
22 P:                                                   [yeah.]
23 D: that i[t  ha]s anything to do with this issue.
24 P:       [.yeah]
25 P: .yes
26     (0.2)
27 P: ye[s.
28 D:   [.hhh good.


*During the early phase of the extract, the doctor turns her back to the
camera and eventually blocks the view of the patient as well.
```

FIGURE 3
Extract 1—Managing disaligning epistemic stances to achieve closer alignment.

ask a question and prefacing it—in short, despite being offered the chance to do it, he seems to treat it as a dispreferred action, and his right to know about it as low.

In her treatment of the patient's idea, the doctor disagrees with it, but also downgrades her epistemic authority by using the word *believe* (lines 14–15). Thus, while the patient treats the doctor as the more knowledgeable party (from whom it is relevant to ask for information) the doctor partially denies this position. By doing so, she manages to simultaneously appeal to her own professional expertise on one hand, and to present the matter as her personal viewpoint on the other.

Interestingly, the patient produces his third turn to the doctor's answer in overlap with its ending, potentially expressing affirmation of what the doctor has just said (see Drew, 2009, on the relationship between overlap onset and affirmation). In his turn, the patient agrees with the doctor's view (that vaccination is not the cause for his symptoms) and recycles the word *believe* as well as the Finnish clitic *-kään*, producing further non-commitment to the idea in a co-operative way (Goodwin, 2013; VISK §839).[2] After establishing this shared evaluation and positioning toward the idea, the doctor repeats her view (lines 21–23), this time upgrading her epistemic stance by using the expression *consider probable* (in comparison to mere believing) and using the expression *anything to do with it*.

In this example, both interlocutors seem to orient primarily to the epistemic stance, and carefully manage the disalignment of their stances. By mitigating his epistemic status and providing his evaluation of the topic at hand as a vague idea in a question format, the patient seems to pre-emptively prepare his stance as susceptive to change. Even though the doctor disagrees with the patient's idea on a factual level, the stance itself is designed so that it disaligns as little as possible. The fact that the patient rushes to confirm the doctor's evaluation of his idea, further works toward alignment between the two. Only after the patient has displayed a stance that abandons the idea, does the doctor strengthen her epistemic stance in declining the idea. By doing so, both participants manage in a stepwise manner to introduce and address a topic that could potentially lead to disalignment, and obtain a shared understanding about it while avoiding the overt disalignment. That is, neither of the interlocutors display a strong stance toward the topic initially but modifies their display of stance in relation to the stance of the other.

Against this background, the second extract exemplifies how, while displaying an epistemic stance toward an idea, patients can also (and potentially primarily) make deontic and affective stances relevant. To achieve this, the patients can employ different turn-design features, such as extreme-case formulations and bodily imitations. Correspondingly, doctors can manage simultaneous and potentially divergent stances in their responses. We unveil the extract in five parts. The first part takes place when the patient has just received a diagnosis. The doctor has initiated the decision-making sequence with a conditional imperative (*now we should start the medication then*). After the patient's preliminary acceptance, the doctor has progressed to describe the dosage and the side-effects. We join the action as the doctor has just mentioned the possible changes in the blood counts and starts to describe a more serious side-effect: intolerance to the medicine (Part 1) (Figure 4).

The doctor describes the potential yet unlikely case of drug allergy and provides information that would help the patient to recognize the side effect. This is done by first comparing the medicine at hand to other medicines and penicillin specifically (lines 3–6), and then by providing information about the rash it might cause (lines 8–13), the rash's potential location (lines 13–15) and the bodily sensations the patient can observe to detect the side effect (lines 17, 19). The depiction is done multimodally, through combination of talk and gesture (pic 3.2 & 3.3). Throughout this segment, the doctor performs one of the basic tasks of the consultation, providing information, thus displaying epistemic expertise. When the doctor vocalizes the word *penicillin* the patient both nods and produces a minimal response (lines 4–5), thus displaying more intense epistemic recognition than earlier during the turn.

The overall activity, suggesting a medication, makes the patient's deontic stance, that is, accepting or rejecting the suggestion, relevant. Furthermore, the doctor's detailed description of symptoms of drug allergy makes it potentially a relevant topic for the patient to address: since the participants have not discussed the patient's drug allergies before, any information considering it would be new information for the doctor, which may, then, be relevant with regard the final choice of medication. This launches the patient to tell about her earlier experience of a penicillin-induced allergic reaction (Part 2) (Figures 5–7).

Epistemic, deontic and affective axes intertwine in the patient's response. Through her multi-unit turn, the patient brings forward her understanding of her existing allergy as potentially relevant for the decision. Both the beginning and the end of the patient's turn depict the relevance of deontic orientation of this turn (see, e.g., Levinson, 2012 on the importance of the beginning and Sacks, 1987 on the endings of turns). The turn starts with the Finnish particle *no* (*well*, line 21), which can be used to produce an action that is not straightforwardly aligning with the previous action (Sorjonen and Vepsäläinen, 2016, 258–265) and thus implicates a need to negotiate before making a decision. Thus, against the sequential projection of a turn that should be doing accepting (or rejecting) the doctor's suggestion, this turn beginning makes it apparent that this alignment is dependent on evaluating the information brought forward by the patient.

The patient informs the doctor that she has a drug allergy to a specific penicillin (line 21–22), and treats it firstly, as something she can evaluate: the described rash symptoms match the patient's experience. The patient works multimodally to make this matching apparent in her response by recycling similar gestures as the doctor does when describing the rash (see lines 13–15 and pic3.2 and pic3.3 for the doctor's gestures and lines 37–40 and pic3.9 for the patient's matching gesture). Secondly, by matching her gestures with those of the doctor, she also manages to build connection between the two epistemic domains: her experience with medications and the side-effects are similar to the ones the doctor has just described (see Goodwin, 2013 on earlier actions as *substrate*). By doing so, the patient positions herself epistemically as capable to identify the allergic rash, showing her recognition and understanding of what the doctor has just described.

The patient ends her turn with an upshot formulation on how she should abstain from using specific type of penicillin, a notion that is being supported by citing another medical professional (see Goffman, 1981, pp. 124–159, on footing and the differentiation of speaker roles).

2 VISK is the online version of the comprehensive Finnish grammar, edited by the Finnish Literature Society.

Part 1

d: >>gaze to P, arms crossed on table->

p: >>gaze to D, hands crossed on lap->

1 D: #.hhhhh (.) >well then% maybe %the< most important thing here (0.3) like

d: %,,,,,,%gaze to papers on table->

#pic3.1

2 D: at the beginning is that maybe% about %ten percent of people are

d: ->%.....%gaze to P->

3 are over sensitive to this medicine.+=exactly like+ one can be

p: +-----nods----+

4 D: to aspirin or +peni[cil]lin+ or some, .hh other

5 P: [mm]

p: +----nods----+

6 D: medicine as well.

7 (0.8)

8 D: .mt and that oversensitivity rash is usually (.) o-

9 D: .hhh+h

p: +nodding->

10 P: yes.

11 D: so it is kind of a+ a bit like well: (.) hives >or<

p: ->+stops nodding

12 >actually< (.) t- >that kind of a bit like< rubella-like

13 +pox-like %ras%#h+ which can #occur, .hhhh (0.2) +like

d: %...%gesturing on chest->

FIGURE 4
Extract 2—Divergent epistemic, affective and deontic stances and their management—Part 1.

Thus, by citing another professional, the patient can produce strong epistemic and deontic statement (*it must be avoided*) while avoiding interfering the professional epistemic domain or displaying a strong personal commitment to the statement. The turn-ending *että* (*so,* line 45) with a continuing intonation also hints that some kind of conclusion should be made from what has been said (Koivisto and Voutilainen, 2016), offering the doctor the chance to evaluate it. Thus, the patient marks her earlier experience as potentially relevant and offers it to the doctor to take a stance on.

The patient's turn also involves elements that make affective stance relevant. By making the affective axis relevant, in addition to explicating her distressing experience, the patient also manages to bring forward her perspective in a professional-lead, institutional situation, where her means for steering the agenda are limited. The patient achieves this by a multimodal gestalt of talk and gesture. The word choices of the story, namely *severe* (line 30), *palm-sized blisters* (line 37–38) and the extreme-case formulation *I thought that I would die* (line 32, Pomerantz, 1986)

Part 2

21 P: +[.hh [okay.]+ [we:ll: (0.2)
p: +-------nodding----------+
22 P: Allergy for Brand A% [I do% have?#
23 D: [yes.
d: %,,,,,,%facing computer->
d: ->%.....%right hand on mouse

#pic3.5

24 D: yes.
25 (.)
26 P: s%o,
d: %click
27 (0.2)%(0.4)
d: %click
28 D: yes.
29 P: it has (.) >the one time< when (.) we noticed it well then it was really
30 P: so severe% th%at%
d: %.....%head turned to patient->
d: %withdraws hand from the mouse->
31 D: yes.=
32 P: =%(oh)% I thou%ght that I will die to this li%ke,#
d: %----nodding----% %nodding->
d: ->%arms crossed on table->

#pic3.6

FIGURE 5
Extract 2—Part 2.

portray her experience with side effects as exceptionally intense and severe. These words are produced as part of multimodal gestalts, where their affective importance is drawn from their precise concordance with matching gestures: the word *palm-sized* with rubbing hands together (line 37, pic3.7 and pic3.8), *blisters* with rubbing the chest (line 38, pic3.9) and *face swell* with gesturing to the face with open palm (line 38, pic 3.10). Through these gestalts the patient manages to upgrade her affective stance. Right after the client has described the size of the blisters, the doctor aligns with the affective stance displayed by the patient, by uttering *oh my* (line 39). This minimal turn allows her to immediately respond to the patient's experience empathetically in institutionally relevant, task-oriented way, which allow the patient to continue her story.

In her turn, the patient has managed to bring forward her epistemic stance (she treats her penicillin allergy as relevant here), deontic stance (should the penicillin allergy prove to be relevant by the doctor, the medicine should be changed) and affective stance (the earlier experience has been very distressing). The following parts of the extract depict how the doctor responds to the patient, first with a

33 P: .hhh[h [like back then (.) back then% well like like (0.5) [I said] to dad
34 D: [okay [yeah. [yes.]
d: ->%nodding finished
35 P: that take photos 'cos I have never had anything like% this
d: %nodding->
36 D: y[es. % [yes.] [yes.]
d: ->%nodding finished
37 P: [.nfft% [so +it] went completely# like (.) [pal]m-sized#
p: +gesturing->(37.40)
#pic3.7 #pic3.8

38 P: kind [of] blisters were# [like and# face swell so then (.)
39 D: [oh my.] [(.yeah .yeah)
#pic3.9 #pic3.10

40 P: we [%took] %like foto's% li[ke] (from the up%per body) back
41 D: [%yes.] [yes.]
d: %------nodding------%
d: %.....%pen in hand, moves->

FIGURE 6
Extract 2—Part 2.

disaligning epistemic stance that also contains an affective dimension (Part 3) (Figure 8).

Overall, the doctor adopts high epistemic status in her stance that contrasts what the patient has said. The doctor first orients to the epistemic axis through providing information. This is initially done by an evaluation: there is a crucial difference between the medicine the patient is allergic to and the one to be prescribed (line 46). The *well* preface marks the contrast between what the patient just said and what the doctor is about the say and the *hän*-clitic marks the information as common knowledge (VISK §1,681). Furthermore, by using the expression "*a drug from a completely different group*" the doctor disaligns with the patient's understanding of the penicillin allergy as being potentially relevant here by negating this. Thus, the doctor builds disalignment with what the patient has said and produces her epistemic authority on the topic. In so doing, the doctor also implicitly disaligns with the patient's deontic stance that this issue should be taken into account in decision making about the medication. An affective orientation is also invested in the expression: the doctor's

42 P: back+ and fort like,%# .h[hh .nfft the doctor said that (.)
p: ->+hands on lap->
43 D: [yes.
d: ->%writing->
#pic3.11

44 P: back then that it should be ↑avoided cos' next time it can be
45 P: even worse s%o(h),=
d: %lifts gaze to P->
d: %stops writing, withdraws hands

FIGURE 7
Extract 2—Part 2.

Part 3
46 D: =%yes. right. .hh %well >this is< a medicine from a completely
d: %gaze to P->
d: ->%arms crossed on table->
47 D: different [group?]
47 P: [yes.]
48 D: a:nd (.) and generally speaking that thing like (0.4) (.) like (.) (Brand-)
49 D: over sensitivity to penicillin is fairly co[mmon] as well.
50 P: [yes.]
((8 lines omitted. The doctor explains that cross allergy between penicillin and the drug at hand does not exist.))
59 D: .hhh (0.3) well of course i- (.) it is in theory possible that(.) that one
60 has like some kind of (.) *eh eh* some kind of a well hereditary
61 D: pred+isposition that one is+ over sensitive to more than one+ drug but
p: +-------nodding--------+
62 D: but propa+bly [that] is not+ the case.
63 P: [mm.]
p: +-----nodding-----+

FIGURE 8
Extract 2—Part 3.

extreme case formulation *'from a completely different group'* can be heard as soothing the patient's expressed worry about the potential side effects.

After elaborating the relationship between the two medicines (omitted) the doctor tells about the possibility of being allergic to multiple medicines (lines 59–62), framing this notion as *theoretical*

and *unlikely*. The patient receives this information with nodding and a response token *mm*, thus aligning with the doctor's more knowledgeable epistemic stance.

The doctor then proceeds to further address the deontic axis of the patient's turn. Interestingly, this treatment of the deontic aspect also works to orient to the patient's affect by displaying immediate interest in the patient's well-being (Figure 9).

In spite of disaligning with the patient's epistemic stance, the doctor aligns with the patient's deontic stance—but with qualifications. Against the background of the unlikeliness of rash, the doctor states how even a *doubt* of rash warrants abstaining from the medication (line 64–65). It is noteworthy that here the doctor formulates the need to abstain from taking the medication as a strict order, enacting her epistemic and deontic status as a medical expert. Thus, while the doctor in essence disaligns with the patient's epistemic and deontic stance (that the patient's history of penicillin allergy should be a contra-indication to the suggested medication), she builds a hypothetical scenario that allows partial alignment. In addition, she provides the patient with necessary information on how to act if side-effects appear, thus aligning with patient's epistemic positioning that she is capable to recognize the allergic rash.

The doctor also implies that despite the patient's incorrect understanding about the medication, her previous experience, and the worry it brings, are being taken into account, thus aligning with the patient's affective stance. Furthermore, by using an extreme-case formulation *do not put a single tablet in your mouth*, the doctor frames this issue as urgent and of high importance. Thus, she orients to a need to negotiate on deontic and epistemic aspects of proper treatment, and resorts to her medical expertise in taking the epistemic and deontic authority and responsibility in regards the result of the negotiation. In addition, the doctor manages the potential discrepancy by aligning with the patient's affective stance, as with her extreme case formulation she conveys her understanding of the severeness of possible side effects that the patient has described in relation to her allergy.

Here it is apparent how the doctor disaligns with some of the patient's stances and aligns with others: she corrects the patient's notion about the potential relevancy of allergy to penicillin but aligns with the patient's intense negative experience, achieving this in institutionally relevant minimal ways. Furthermore, regarding the deontics, she manages to both align and disalign with the patient's hesitative stance on whether to accept the suggested medication or not. This is achieved by first providing information and reasoning why the patient's idea about the penicillin allergy is not relevant in this case and then providing strong description of how to handle the conditional and unlikely event of getting allergic rash, which does align with the patient's affective stance.

Despite the doctor having addressed all three stances, the patient expands her experience-telling and provides yet another epistemic evaluation of her previous rash (Part 5) (Figure 10).

The doctor has ended her turn with information about the timing of the occurrence of the side-effects (lines 68–70, part 4). The patient picks up on this (lines 72–82), displaying again the resemblance of her experience and what the doctor has described. At this point it is

```
Part 4
64 D: well of course if you then doubt that you are +getting a rash
   p:                                                +nods slowly->
65 D: +then you do not put a single tablet in your mouth.
   p:>+nod finished
66    (.)
67 P: yes.+
   p:     +nodding->
68 D:.hh[h (.) it always comes with+ a delay (.) +[if  ]it comes at all
69 P:   [yes.                                    +[yes.]
   p:                              ->+nod. stopped +nodding->
70 D: so yo- (.) it+ comes within a few weeks [of ti]me
71 P:                                         [yes. ]
   p:            ->+nodding stopped
```

FIGURE 9
Extract 2—Part 4.

ambiguous whether the patient treats her experience still epistemically and deontically relevant regarding the decision-making: that is, if she continues her telling to show that she knows what a delayed effect might be and her ability to recognize one if necessary (thus aligning with the decision) or to make explicit that similar to the medicine to be prescribed also the earlier one caused delayed side-effects (thus resisting the decision).

Whether the patient is aligning or disaligning, the doctor expresses her understanding about what the patient is telling early on during the patient's turn. As the patient is approaching the point in her story where she has taken the whole treatment, the doctor produces a potential end for the patient's story (lines 80–81, note also that on line 77, she starts a yes-no question, which potentially would have carried the same pre-assumption). The doctor's turn is built as a continuation to the patient's turn (*I took the last pill on Saturday evening and*…) producing a grammatically compatible, collaborative completion to it (… *and then it came*). By this collaborative completion, the doctor shows that she not only understands what the patient is saying, but

```
Part 5
72 P: it was also back then (.) it was a course of antibiotics I don't [remember]
73 D:                                                                  [   mm.  ]
74 P: why [I ate] that course of antibiotics like it was .hhhhh one week's
75 D:     [yes. ]
76 P: course I think.
77 D: mm=[did it com- ]
78 P:    [and  the  la]st tablet I took like (0.2) on Friday: (.) evening
79 D: yeah.
80 P: and (.) on sa[TURDAY?   ]
81 D:              [after% that] it came.
   d:                   %nodding->
82 P: on Sa[turd]ay [so  ] (.) I [wonder]ed like
83 D:      [yeah.]  [yes.]       [ .yes ]
84 P: [like]% did+ the medicine protect [it like+ or]
   p:            +---------gesturing------------+shrugs shoulders->
85 D:  [mm.]                           [no.  it  is] it is
   d:     ->%nodding stopped
86 d: by na+ture this kind of [delay-] delayed over sensitivity reaction
87 P:                         [↑yes. ]
   p:    ->+hands on lap>>
88 D: [that ] appears there like, .hh [typi]cally after two to
89 P: [↑joo.]                        [yes.]
90 L: four weeks +[with this] medication if it is about to co+me [at all.]
91 P:            +[  right. ]                                    [ yes.  ] okay.
   p:            +---nodding--+                                +nodding->
 92 (0.4)+(0.3)
   p: ->+nodding finished
93 D: .mt %(.) but% it is like +kind of an important th+ing to know.
```

FIGURE 10
Extract 2—Part 5.

also that she *independently* knows this, once again producing epistemic authority.

After this, the patient brings forward another idea: the medication might have protected her from the rash (lines 82–84). Once again, the question-format enables her to present the idea without a strong commitment to it, supported by shrugging the shoulders. In her response, the doctor first orients to the epistemic axis by, again, disproving the idea (lines 85–90), followed by orientation to deontic axis, by repeating her earlier instructions (95–100), and finally to affective axis, by highlighting the immediacy of reacting to potential side-effects (*immediately*, lines 95 and 100). After this, the participants proceed to briefly discuss the interaction between the prescribed medicine and other medicines, before the doctor prints out the prescription.

Discussion

In this article we have illustrated (1) how participants orient to multiple, even divergent stances as relevant at the same time and even in the same turn of talk, and (2) manage them with multimodal means to achieve closer alignment. In discussing patients' ideas about their symptoms or future treatment, the participants orient to alternating between three axes of stance: deontic, epistemic and affective. They can emphasize the relevance of one over another, as in Example 1 where the epistemic stance was oriented to as focal, or they can address potential disalignment between some of the stances by orienting to the third, aligned one, as in Example 2 where the doctor softened the deontic and epistemic disalignment with the help of displaying aligning affective stance with the patient. The role of context obviously restricts or emphasizes the relevance of particular axes of stance, as in the clinical consultation in focus here, the institutional tasks of finding diagnosis and proper treatment for the patients' ailment forms the main aim of the encounter, this way foregrounding the relevance of deontic and epistemic stances. As the epistemic and deontic statuses of the doctor are oriented to as higher, affective stance may afford more leeway for the patients to bring forward their own topics and concerns within the rather structured agenda of medical consultations. By evoking their own experiential knowledge and displaying affective stance they can interfere into the medical doctor-driven activity to introduce their own projects and manage their potentially disaligning stances concerning treatment decisions (see example 2).

We have also illustrated how the management of stances is multimodal. Epistemic and deontic stance may be conveyed for example through reported speech, and affective stance may be displayed both verbally, for example through extreme case formulations, and non-verbally, for example when enacting emotionally strenuous situations through gestures. The same multimodal turns can also be used to convey multiple stances at the same time, as in Example 2 where the patient imitated the doctor's gestures in displaying her epistemic stance (in giving information about her previous rash) and her affective stance (by combining her extreme case formulations with the imitative gestures) (c.f. Mondada, 2018, on how multimodal actions may enable multiple temporal progressions in interaction).

Is there then some primacy between the management of stances? For example, should the stance(s) where the disalignment stands have a priority in the management process? This is simply too early to claim, as our data provides possibilities for both interpretation that this is the case (in extract 2, the doctor starts her multi-unit turn with epistemic orientation) and that this is not the case (already during the troubles-telling, the doctor uses *oh my* to do empathy). However, our data shows that being able to align on some axis may work as a lever to shift the disaligning stances on other axes toward closer alignment.

It is noteworthy that the management of stances entails subtle negotiation, turn by turn, where the displays of stance and also the balance of the different axes of stance are slightly modified, optimally resulting in reasonably sufficient intersubjective understanding and alignment which are necessary in order to proceed from one activity to another. In displays of stance, a certain cautiousness is observable: the original stance may be presented as open to change (example 1) or in a delicate way (example 2). This way, management of the simultaneous axes of stance appears to contribute both to maintaining the relationship of the participants and the progressivity of the consultation.

Our starting points have been the theoretical findings of Du Bois (2007), who has highlighted the intersubjective nature of stance and the interlocutors' orientation to the alignment of stances, as well as the empirical work by Küttner (2019), Tai et al. (2022) and Sakita (2013, 2017), who have emphasized the negotiable and repairable nature of stance. Building on their observations and our own empirical work (see also Ruusuvuori and Peräkylä, 2009; Logren et al., 2019; Logren et al., 2020), we argue that in addition of straightforward *stance-taking,* a more flexible and situated *management of stance* was apparent in our cases. That is, in each action the participants took a stance in relation to one or more axes, but did it in ways that allowed the participants to adjust their stances and thus form alignment action-by-action (for example by expressing openness to change, delicacy or hypothetical scenarios). Furthermore, we have illustrated how the epistemic, deontic and affective axes of stance not only intertwine in interaction, but they have a crucial relationship with each other, into which the interlocutors observably orient to as relevant and consequential in relation to how their ongoing project can proceed. Nevertheless, the three axes are not, and do not need to be constantly parallel—as we have shown, their divergence *per se* seems not to be problematic for the interlocutors. Instead, divergence can actually work as a way to handle the disaligment of some of the stances: alignment in one stance seems to function to mitigate or even repair disalignment in the others.

We want to raise three points for methodological reflection. First, the examples we have provided here come from different types of sequential context: namely, asking a question at a transition between phases of the overall structure of the encounter (Example 1) and accepting/declining a suggestion during the decision-making phase (Example 2). Participants orient to achieving different tasks in different phases of a consultation which may shape both if and how patients bring forward different stances and how professionals respond to them. More robust, collection-based analysis of divergent stances in specific sequential environments followed by a systematic comparison of different contexts should provide steps for future theory building.

Second, our data are from a specific institutional context and as has been noted, this shapes the management of both epistemics, deontics and affect (Drew and Heritage, 1992; Heritage and

Clayman, 2010; Jefferson and Lee, 1992; Robinson and Stivers, 2001; Ruusuvuori, 2005; Stivers, 2002). It has been suggested that the central task of institutional CA is describing *possible* social actions and their sequential and institutional conditions and the generalizability of these findings stem from comparison of the findings (Peräkylä, 2004). Thus, we do not claim generalizability of our findings but present potential questions to be asked from data from different context. As the management of divergent stances might be subordinate to the institutional task at hand, a question arises: how do interactants orient to different institutional relevancies when managing divergent stances? For instance, in psychotherapeutic contexts, the institutional task entails working with the patient's emotions, and therefore the management of divergent stances can actualize in a very different form than in medical consultations. In addition, analyzing mundane conversations would provide even more insight on how divergent stances are managed without the institutional frameworks where specific deontic, epistemic or affective statuses may be inscribed in the institutional task. In short, more diverse data are needed to better understand the phenomenon.

And third, as the second extract exemplified, multimodal management of divergent stances can be messy, scattered and take a lot of time. Earlier research on, for example deontics, has focused on reasonably straightforward three-turn-structures (with clear benefits, such as producing robust theoretical formulations). However, when more than one stance is at stake, it seems that the action can disintegrate: resembling the tentacles of an octopus or a mycelium, one line of action can take one direction while others continue another way, just to compound at some unexpected moment. Studying complex phenomena such as the management of multiple simultaneous stances might then benefit from analytic strategies that expand beyond the sequence and follow the topical progression beyond the boundaries of an action sequence.

We conclude by stating that people do not just take a stance and stick with it: rather it seems that people modify their stances slightly as the interaction progresses, taking into account the stances displayed by their interlocutors. In each turn of talk, epistemic, deontic and affective stances are laminated, various stances are taken in the same turn, and the individual stances of the interlocutors are step by step shifted closer to each other—from disalignment to alignment. Therefore, interlocutors can eventually achieve at least partial alignment which they orient to as sufficient to enable them to proceed in the ongoing activity or task at hand without overt conflict or rupture in their relationship. Management of stances is thus not just a structural feature of interaction but may crucially influence the relationships of the interlocutors. It is observably relevant for the participants both in its local sequential context as well as in terms of the tasks they pursue together.

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.

Ethics statement

The studies involving humans were approved by the local Ethics Committee for Medical Research in Pirkanmaa region, Finland (reference code R21057). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

AL: Writing – original draft, Writing – review & editing. SI: Writing – original draft, Writing – review & editing. JR: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study has been supported by the Strategic Research Council (SRC) established within the Academy of Finland (project numbers 31213358415 and 31213584181).

Acknowledgments

Professors Hanna Kuusisto (UEF) and Reetta Kälviäinen (UEF) provided an invaluable contribution to the data collection of which the authors want to express their deepest gratitude.

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/fpsyg.2024.1436677/full#supplementary-material

References

Antaki, C., and Kent, A. (2012). Telling people what to do (and, sometimes, why): contingency, entitlement and explanation in staff requests to adults with intellectual impairments. *J. Pragmat.* 44, 876–889. doi: 10.1016/j.pragma.2012.03.014

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Clift, R. (2021). Embodiment in dissent: the eye roll as an interactional practice. *Res. Lang. Soc. Interact.* 54, 261–276. doi: 10.1080/08351813.2021.1936858

Curl, T. S., and Drew, P. (2008). Contingency and action: a comparison of two forms of requesting. *Res. Lang. Soc. Interact.* 41, 129–153. doi: 10.1080/08351810802028613

Drew, P. (2009). "Quit talking while I'm interrupting: a comparison between positions of overlap onset in conversation" in Talk in interaction: comparative dimensions. eds. M. Haakana, M. Laakso and J. Lindström (Helsinki: Finnish Literature Society), 70–93.

Drew, P., and Heritage, J. (1992). "Analyzing talk at work: an introduction" in Talk at work. Interaction in institutional settings. eds. P. Drew and J. Heritage (Cambridge: Cambridge University Press), 3–65.

Du Bois, J. W. (2007). "The Stance Triangle" in Stancetaking in discourse. ed. R. Englebretson (Amsterdam, Philadelphia: John Benjamins Publishing Company), 139–182.

Du Bois, J. W., and Kärkkäinen, E. (2012). Taking a stance on emotion: affect, sequence, and intersubjectivity in dialogic interaction. *Text Talk* 32, 433–451. doi: 10.1515/text-2012-0021

Ekström, M., and Stevanovic, M. (2023). Conversation analysis and power: examining the descendants and antecedents of social action. *Front. Sociol.* 8:1196672. doi: 10.3389/fsoc.2023.1196672

Gill, V. T. (1998). Doing attributions in medical interaction: patients' explanations for illness and doctors' responses. *Soc. Psychol. Q.* 61:342. doi: 10.2307/2787034

Gill, V. T., and Maynard, D. W. (2006). "Explaining illness: patients' proposals and physicians' responses" in Communication in medical care. Interaction between primary care physicians and patients. eds. J. Heritage and D. W. Maynard (Cambridge: Cambridge University Press), 115–150.

Goffman, E. (1981). Forms of talk. (Philadelphia: University of Pennsylvania Press).

Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *J. Pragmat.* 46, 8–23. doi: 10.1016/j.pragma.2012.09.003

Goodwin, M., Cekaite, A., and Goodwin, C. (2012). "Emotion as stance" in Emotion in interaction. eds. M.-L. Sorjonen and A. Peräkylä (Oxford: Oxford University Press), 16–41.

Halkowski, T. (2006). "Realizing the illness: patients' narratives of symptom discovery" in Communication in medical care. (Cambridge: Cambridge University Press), 86–114.

Heritage, J. (1984). "A change-of-state token and aspects of its sequential placement" in Structures of social action: studies in conversation analysis. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 299–345.

Heritage, J. (2009). "Questioning in medicine" in "Why do you ask?" The function of questions in institutional discourse. eds. A. F. Freed and S. Ehrlich (Oxford: Oxford University Press), 42–68.

Heritage, J. (2012). "Epistemics in conversation" in The handbook of conversation analysis. eds. J. Sidnell and T. Stivers (Chichchester: Wiley), 370–394.

Heritage, J. (2013). Action formation and its epistemic (and other) backgrounds. *Discourse Stud.* 15, 551–578. doi: 10.1177/1461445613501449

Heritage, J., and Clayman, S. E. (2010). Talk in action: interactions, identities, and institutions. Chichester: Wiley-Blackwell.

Heritage, J., and Raymond, G. (2005). The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. *Soc. Psychol. Q.* 68, 15–38. doi: 10.1177/019027250506800103

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction" in *Conversation Analysis*. Studies from the first generation (Amsterdam, Philadelphia: John Benjamins Publishing Company), 13–31.

Jefferson, G., and Lee, J. R. E. (1992). "The rejection of advice: managing the problematic convergence of a 'troubles-telling' and a 'service encounter" in Talk at work. eds. P. Drew and J. Heritage (Cambridge: Cambridge University Press), 521–548.

Kaukomaa, T., Peräkylä, A., and Ruusuvuori, J. (2013). Turn-opening smiles: facial expression constructing emotional transition in conversation. *J. Pragmat.* 55, 21–42. doi: 10.1016/j.pragma.2013.05.006

Koivisto, A., and Voutilainen, L. (2016). Responding to what is left implicit: psychotherapists' formulations and understanding checks after clients' turn-final *Että* ("that/so"). *Res. Lang. Soc. Interact.* 49, 238–257. doi: 10.1080/08351813.2016.1198195

Koskinen, E., and Stevanovic, M. (2022). Epistemic calibration. *Pragmatics* 32, 354–380. doi: 10.1075/prag.20036.kos

Küttner, U.-A. (2019). At the intersection of stance-management and repair. *Gesprächsforschung: Online-Zeitschrift Zur Verbalen Interaktion* 20, 115–156.

Land, V., Parry, R., and Seymour, J. (2017). Communication practices that encourage and constrain shared decision making in health-care encounters: systematic review of conversation analytic research. *Health Expect.* 20, 1228–1247. doi: 10.1111/hex.12557

Levinson, S. C. (2012). "Action formation and ascription" in The handbook of conversation analysis. eds. J. Sidnell and T. Stivers (Chichchester: Wiley), 101–130.

Local, J., and Walker, G. (2008). Stance and affect in conversation: on the interplay of sequential and phonetic resources. *Text Talk* 28, 723–747. doi: 10.1515/TEXT.2008.037

Logren, A. (2019). The management of experience as a platform for social influence in health promotion groups. (Doctoral dissertation): Tampere University Dissertations.

Logren, A., Ruusuvuori, J., and Laitinen, J. (2019). Peer responses to self-disclosures in group counseling. *Text Talk* 39, 613–647. doi: 10.1515/text-2019-2042

Logren, A., Ruusuvuori, J., and Laitinen, J. (2020). Stories of change: comparative time-framed experience telling in health promotion group discussions. *Qual. Health Res.* 30, 279–292. doi: 10.1177/1049732319877858

Mondada, L. (2001). Conventions for multimodal transcription. Available at: https://www.lorenzamondada.net/multimodal-transcription (Accessed September 5, 2024).

Mondada, L. (2018). Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51, 85–106. doi: 10.1080/08351813.2018.1413878

Peräkylä, A. (1998). Authority and accountability: the delivery of diagnosis in primary health care. *Soc. Psychol. Q.* 61, 301–320. doi: 10.2307/2787032

Peräkylä, A. (2004). Reliability and validity in research based on naturally occurring social interaction. *Qualitative research. Theory, method and practice.* ed. D. Silverman 283–304.

Peräkylä, A., and Silverman, D. (1991). Owning experience: describing the experience of other persons. *Text* 11, 441–480. doi: 10.1515/text.1.1991.11.3.441

Petraki, E., and Clark, S. (2016). Affiliating through agreements: the context of antenatal consultations. *Austral. J. Ling.* 36, 273–289. doi: 10.1080/07268602.2015.1121535

Pomerantz, A. (1984). "Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes" in Structures of social action. Studies in conversation analysis. eds. J. M. Atkinson and J. Heritage (Cambridge: Cambridge University Press), 57–101.

Pomerantz, A. (1986). Extreme case formulations: a way of legitimizing claims. *Hum. Stud.* 9, 219–229. doi: 10.1007/BF00148128

Robinson, D., and Stivers, T. (2001). Achieving activity transitions in physician-patient encounters: from history taking to physical examination. *Hum. Commun. Res.* 27, 253–298. doi: 10.1111/j.1468-2958.2001.tb00782.x

Ruusuvuori, J. (2000). Control in the medical consultation. Practices of giving and receiving the reason for the visit in primary care consultations. (Doctoral dissertation). *Acta Electronica Universitatis Tamperensis* 16.

Ruusuvuori, J. (2005). "Empathy" and "sympathy" in action: attending to patients' troubles in Finnish homeopathic and general practice consultations. *Soc. Psychol. Q.* 68, 204–222. doi: 10.1177/019027250506800302

Ruusuvuori, J. (2007). Managing affect: integration of empathy and problem-solving in health care encounters. *Discourse Stud.* 9, 597–622. doi: 10.1177/1461445607081269

Ruusuvuori, J., and Peräkylä, A. (2009). Facial and verbal expressions in assessing stories and topics. *Res. Lang. Soc. Interact.* 42, 377–394. doi: 10.1080/08351810903296499

Sacks, H. (1987). "On the preferences for agreement and contiguity in sequences in conversation" in Talk and social organisation. eds. G. Button and J. R. Lee (Bristol: Multilingual Matters), 54–69.

Sacks, H. (1992). Lectures on conversation. Oxford: Basil Blackwell.

Sakita, T. I. (2013). Discourse markers as stance markers. *Pragmat. Cogn.* 21, 81–116. doi: 10.1075/pc.21.1.04sak

Sakita, T. I. (2017). Stance management in oral narrative. *Funct. Lang.* 24, 65–93. doi: 10.1075/fol.24.1.04sak

Schegloff, E. A. (1987). Analyzing single episodes of interaction: an exercise in conversation analysis. *Soc. Psychol. Q.* 50, 101–114. doi: 10.2307/2786745

Sorjonen, M.-L., and Vepsäläinen, H. (2016). The Finnish particle no. *NU / NÅ: A Family of Discourse Markers Across the Languages of Europe and Beyond.* eds. P. Auer and Y. Maschler Berlin, Boston: De Gruyter. 243–280. doi: 10.1515/9783110348989-008

Stevanovic, M. (2018). Social deontics: a nano-level approach to human power play. *J. Theory Soc. Behav.* 48, 369–389. doi: 10.1111/jtsb.12175

Stevanovic, M., and Peräkylä, A. (2012). Deontic authority in interaction: the right to announce, propose, and decide. *Res. Lang. Soc. Interact.* 45, 297–321. doi: 10.1080/08351813.2012.699260

Stevanovic, M., and Peräkylä, A. (2014). Three orders in the organization of human action: on the interface between knowledge, power, and emotion in interaction and social relations. *Lang. Soc.* 43, 185–207. doi: 10.1017/S0047404514000037

Stivers, T. (2002). Participating in decisions about treatment: overt parent pressure for antibiotic medication in pediatric encounters. *Soc. Sci. Med.* 54, 1111–1130. doi: 10.1016/S0277-9536(01)00085-5

Stivers, T. (2005). Parent resistance to physicians' treatment recommendations: one resource for initiating a negotiation of the treatment decision. *Health Commun.* 18, 41–74. doi: 10.1207/s15327027hc1801_3

Stivers, T. (2006). "Treatment decisions: negotiations between doctors and patients in acute care encounters" in Communication in medical care. Interaction between primary care physicians and patients. eds. J. Heritage and D. W. Maynard (Cambridge: Cambridge University Press), 279–312.

Stivers, T., Mondada, L., and Steensig, J. (2011). "Knowledge, morality and affiliation in social interaction" in The morality of knowledge in conversation. eds. T. Stivers, L. Mondada and J. Steensig (Cambridge: Cambridge University Press), 3–24.

Tai, A., Nguyen, H., and Vicentini, C. (2022). Interactional practices to manage epistemic stances in online searches during a computer-mediated conversation-for-learning. *TESOL Context* 30, 39–63. doi: 10.21153/tesol2022vol30no2art1517

Toerien, M. (2018). Deferring the decision point: treatment assertions in neurology outpatient consultations. *Health Commun.* 33, 1355–1365. doi: 10.1080/10410236.2017.1350912

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 08 April 2025
DOI 10.3389/fpsyg.2024.1453168

Check for updates

OPEN ACCESS

EDITED BY
Elisabeth Zima,
University of Freiburg, Germany

REVIEWED BY
Claudia Lehmann,
University of Potsdam, Germany
Ali Alsaawi,
Majmaah University, Saudi Arabia

*CORRESPONDENCE
Sabina Tabacaru
✉ sabina.tabacaru@univ-paris8.fr

RECEIVED 22 June 2024
ACCEPTED 23 September 2024
PUBLISHED 08 April 2025

CITATION
Tabacaru S (2025) "That was really a productive segment, wasn't it?" Nonverbal markers of humor in the American presidential debates of 2016 and 2020. *Front. Psychol.* 15:1453168. doi: 10.3389/fpsyg.2024.1453168

COPYRIGHT
© 2025 Tabacaru. 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.

"That was really a productive segment, wasn't it?" Nonverbal markers of humor in the American presidential debates of 2016 and 2020

Sabina Tabacaru[1,2]*

[1]Département d'Etudes des Pays Anglophones, UFR LLCER-LEA, Université Paris 8, Saint-Denis, France, [2]UR TransCrit, Saint-Denis, France

These past few years have marked a growing interest in multimodality, interaction and eye-gaze in the interpretation and understanding of discourse. Eye-gaze, for example, plays a central role in face-to-face interaction and stance taking because it helps discourse participants coordinate with each other. Such visual markers help the interlocutors/audience to intersubjectively connect to the same common ground on which they construe their meanings. The case of humor has also received more attention from a multimodal perspective since it follows the same patterns of meaning construction and coordination. Elements that are salient to the humorous interpretation will be emphasized using either prosodic cues or visual markers, such as facial expressions and head movements. In this paper, we explore the use of such nonverbal discourse markers with the use of humor in the American presidential debates of 2016 and 2020, analyzing their role on the humorous stance.

KEYWORDS

humor, sarcasm, stance, multimodality, nonverbal, political debates

Introduction

Humor holds different functions in conversation (Hay, 2000), and it has been a point of interest for linguists, psychologists, anthropologists, among others. As a research path, it provides complex analyses of discourse since speakers need to make their intentions salient to the other participants. As opposed to other types of analyses of discourse, humor relies on the element of incongruity, where meanings overlap, giving rise to different effects that the speakers need to make obvious to their interlocutors. Incongruity thus implies changing between different meanings, or frames, in order to switch the viewpoint on which the interpretation relies. I align here with frameworks of humor in terms of a layered mental space configuration, as has been explained elsewhere (Brône, 2008) in which meaning is formed as a pretense space on top of a discourse space (see also Tabacaru, 2019 or De Vries et al., 2021).

Moreover, humor has been linked to persuasion (Li and Pryor, 2020), particularly in political arguments used for different purposes (Feldman, 2024), which will also be the case here. For example, Li and Pryor (2020) find that, in oral arguments, there is a positive reaction to the people who cause laughter as they win more votes from justices in the Supreme Court. As such, humor would have an effect on the power of persuasion.

The aim of the present paper is to look at semiotic markers of humor in interaction in order to analyze the role they have in (humorous) meaning construction in the specific context

of political debates, where participants need to argue in favor of their position, as opposed to the one presented by their adversary. Hence, the viewpoints presented are different, contradictory, and each participant needs to persuade an audience that will cast a vote. The understanding of different viewpoints (Dancygier and Vandelanotte, 2016) is central to this type of discourse in order for interlocutors to make the appropriate responses, be their humorous or not. This echoes the theory of mind (Baron-Cohen, 1995; Tomasello, 1999) because each participant needs to understand their opponent's position as well as the complex meanings created through implicature (Grice, 1975).

At the heart of the debate, the emphasis falls on the question of stance, which is understood, according to Du Bois (2007: 139) as "the power to assign value to objects of interest, to position social actors with respect to those objects, to calibrate alignment between stancetakers, and to invoke presupposed systems of sociocultural value." Speakers thus have to position themselves in relation to their opponent and certain topics that are the subject of the debate. It is thus argued that humorous meaning plays a central role in the way stancetakers evaluate their opponent's discourse, by directly criticizing or by switching the viewpoint to a negative one.

In the following, closer attention is paid to interaction and multimodality in discourse, from the perspective of humorous communication.

Humor and interaction

While traditional approaches to humor (Raskin, 1985) paid attention to the semantic and/or pragmatic background that helps create jokes, more recent studies emphasize the role interaction plays in these types of exchanges (Priego-Valverde, 2009; Feyaerts et al., 2015; Brône and Oben, 2013; Feyaerts, 2013a; Feyaerts, 2013b; Feyaerts and Oben, 2014; Tabacaru, 2019, among others). From this perspective, both speaker and interlocutor have to be taken into consideration in the way these meanings are created and understood. Such studies focus on questions of intersubjectivity and common ground, which are fundamental in meaning coordination among speakers. A shared common ground (Clark, 1996), for example, allows the speaker to know which processes their interlocutor will go through in order to understand their message as humorous (see, for example, Yus, 2003, who speaks about the search for relevance). Drawing from shared background assumptions, the speakers are able to coordinate with each other's meanings and behaviors, using the context and discourse at hand. As Geeraerts (2021: 24) puts it: "Ideally, if A and B are paying attention to Z, the common ground is maximal if both A and B believe that the other focuses on Z, and if A and B both believe that the other is aware of the fact that they focus on Z." The speakers thus make references to the same common ground that *they believe* they share with the others. Implied meanings are thus to be understood against a common ground that discourse participants share with each other and that they enrich in conversation.

In order to illustrate this, consider the following example which comes from the present corpus (Trump versus Biden in 2020) and represents an example of humor in interaction. As the debate and the presidential campaign take place during the Covid-19 pandemic, the moderator talks about regulations regarding crowds, stating that the two candidates have different approaches, i.e., Trump prefers to campaign in large crowds, whereas Biden organizes smaller events. The humor is construed by Trump in his last reply, building on what the moderator said:

(1) Wallace: I want to ask you both about one last subject because your different approaches has even affected the way that you have campaigned. President Trump, you're holding large rallies with crowds packed together, thousands of people.
Trump: Outside.
Wallace: Outside. Yes, sir. Agreed. Vice President Biden, you are holding much smaller events with---
Trump: Because nobody will show up.
[moderator laughs][1]

[1]The video is available here: https://www.youtube.com/watch?v=CweqW7Pzxz8 (the example starts at 34m50s).

This represents an example of *dialogicality*, which, as defined by Du Bois (2007: 140) is when "a stancetaker's words derive from, and further engage with, the words of those who have spoken before." The viewpoint is shifted as, at first, the focus of attention is on the crowds that gather to see the candidates, and the moderator here uses the adjectives *large* and *smaller* to describe each candidate's campaign. Of course, the larger crowds would mean more risk of contracting Covid-19, but Trump emphasizes that he holds these rallies outside. This implies that he respects the regulations and minimizes the risks taken. When the word *smaller* is used, which highlights the comparison to his *large* rallies, Trump adds a different perspective: he can organize large events because people want to see him whereas Biden cannot organize large events as nobody will attend these. This is also a case of reasoning (if P, then Q), with different implied meanings:

If P	then Q
If people were interested in Biden's campaign	→ he would also hold larger rallies
If people come to Trump's rallies	→ they are (only) interested in Trump's campaign

The shift of focus from him to his opponent also shows Trump's understanding of the different implicatures between *larger* and *smaller crowds* in the context of the pandemic. In order to shift the viewpoint in the given context and reverse the negative implication from him to Biden, he needs "deep cognitive coordination" (Verhagen, 2005: 4) with the other speaker, as well as the audience watching the debate.[1]

As further noted by Verhagen (2005: 7-8), "a speaker […] is committed to the assumption that her utterance is in principle interpretable by someone else sharing the knowledge of certain conventions." In this case, Trump knows very well the audience can access the same negative implicatures that he did regarding his campaign and the larger crowds attending his rallies. This is made clear by Verhagen (2005: 12): "The inferential load of utterances is crucially involved in the way they relate to each other in connected discourse. Discourse consists of chains of inferential steps, including the possibility of rejecting one or more steps and 'changing course'." This is what happens here, as Trump needs to reject the negative implicatures and put the focus on his opponent rather than his own behavior and campaign. Even though the moderator never specifically says that this comparison is negative, this is interpreted as "negative evaluation" (Hunston, 2007: 41) by Trump, who then has to change the viewpoint and the course of the debate.

The use of the modal *will* is linked to certainty, as, from this point of view, there is no doubt that people do not want to attend Biden's rallies. This fits the category of generalization (see Zhang, 1998, and Scheibman, 2007), which represents a comment on Biden's entire campaign. From Trump's point of view, their Covid-19 behaviors are divergent between

1 This also echoes Langacker's (1987) construal relationship and figure-ground reversal.

the two candidates, but for different reasons. This creates humor in the way the implicatures are created as it has to be interpreted against a shared common ground: in this case, the Covid-19 pandemic and the regulations surrounding it. This example, as many others from the corpus, can be seen as sarcastic, which is represented as a pretense (Clark and Gerrig, 1984; Barnden, 2017) with "clearer markers/cues and a clear target" (Attardo, 2000, 795). Even against Grice's cooperative principle (for example, the maxim of quality), the incongruity between what is real and what is pretense is clear: Trump targets his opponent by creating a pretense space in which people would not attend his rallies because they are not interested in his campaign.

In face-to-face interactions, as is the case here, participants also make their humorous implicatures clear to the interlocutors/audience, through their verbal and non-verbal behavior (see De Vries et al., 2021). Multimodality thus plays a central role in the way speakers coordinate meanings with each other.

Multimodal markers of humor

There is an increasing focus on the way the body is used in interaction, as a marker of stance (see, for example, Bateman et al., 2017; Feyaerts et al., 2017, but also De Vries et al., 2021 specifically on ironic stance). These bodily resources include different elements, such as head tilts and shakes (Jehoul et al., 2017), nods (Mondada, 2009), shrugs (Jehoul et al., 2017), raised eyebrows (Feyaerts et al., 2022), frowning (Li, 2021), eye-gaze (Brône et al., 2017; De Vries et al., 2021 or De Vries et al., 2024), but also hand gestures (see Mondada, 2009; Jehoul et al., 2017). They all have different functions in interactions (for an overview, see Andries et al., 2023), ranging from an emotional perspective (such as surprise; see Ekman and Wallace, 2003) to negative/positive stance.

Some of these elements have been linked to humor production: head movements (Tabacaru, 2019 or De Vries et al., 2021), raised eyebrows and frowning (Tabacaru and Lemmens, 2014). Other non-verbal elements used with humor can include prosodic markers (Bertrand and Priego-Valverde, 2011; Attardo et al., 2013), but also smiling (Gironzetti, 2017) or laughter (Attardo, 2003).

The question is, then, to see the role these elements play with humor. In corpus studies conducted on non-spontaneous uses of humor, Tabacaru and Lemmens (2014) and, later, Tabacaru (2019) explain the role of *gestural triggers* for the shift from a serious meaning to a humorous/pretense meaning. In these contexts, speakers used raised eyebrows with salient elements for the humorous interpretation. The experiment was repeated with more spontaneous uses of humor for the French political debates prior to the presidential elections of 2017 (Tabacaru, 2020). The study shows that even in these types of interactions, which do not aim at being humorous (as opposed to television series and shows, stand-up comedy, etc.), speakers will make use of non-verbal elements in order to make their humorous message understood. In a study investigating eye-gaze in spontaneous interactions, De Vries et al. (2024) found that, in internal teases, the target of the teasing was focused on both verbally and visually, which then highlights the role these semiotic resources have in such settings.

Furthermore, Gironzetti (2022: 15), talks about different types of discourse markers that are used to signal "the speaker's intentions by conveying a metamessage about how a certain utterance should be interpreted." Or, as noted by Verhagen (2005: 22):

> Linguistic expressions are primarily cues for making inferences, and understanding does not primarily consist in decoding the precise content of the expressions, but in making inferences that lead to adequate next (cognitive, conversational, behavioral) moves.

We can consider that speakers make assumptions and draw conclusions from their interlocutors' (both verbal and non-verbal) behavior as discourse unfolds. As such, research should focus more on the multimodal side of discourse (Norris, 2004), as interactions include elements that are essential to the understanding of the (humorous) message. Speakers have to interpret their interlocutors' non-verbal behavior, be it a gaze, facial expressions, body posture, etc., and respond to it accordingly.

As Stubbs (1986: 1; quoted in Englebretson, 2007:70) points out:

> whenever speakers (or writers) say anything, they encode their point of view towards it … The expression of such speakers' attitudes is pervasive in all uses of language. All sentences encode such a point of view, … and the description of the markers of such points of view and their meanings should therefore be a central topic for linguistics.

Englebretson (2007: 70) further notes that "every utterance enacts a stance," hence, for the purpose of the present paper, the focus will fall on the humorous stance that is achieved in the corpus, which will be presented in the section below. As noted by De Vries et al. (2021: 37), it would make sense that such resources are used at specific times if they signal ironic intent. This is also the point made by Tabacaru (2019) where humor was made prominent through some sort of non-verbal element used by the speaker.

Corpus and method

The examples presented here come from two American presidential debates: the first presidential debate between Donald Trump and Hillary Clinton in 2016 (@NBC News on YouTube), of a duration of 1h35m, and the first presidential debate between Donald Trump and Joe Biden in 2020 (©The Telegraph on YouTube), of a duration of 1h39m. Transcripts available online[2] for the debates were also used, which consisted of more than 37.000 words. The settings of such debates are not as rich in humorous exchanges as television series (Brône, 2008, for example) or stand-up comedy shows, which are meant to be humorous by nature. But, as the debates present opposing candidates who need to surpass their adversary's speech and persuade people to vote for them, humor will be used; such interactions also provide non-staged uses of humor, as opposed to other types of discourses mentioned before. They are also rich in interactional cues provided by the speakers. Hence, other meanings can emerge, because at the core of the debate is the wish to persuade people to vote for a certain candidate (Cobb and Kuklinski, 1997). For both these debates, the audience is reminded (either before or throughout the debate) that they are expected to remain silent, but there are

2 The commission on presidential debates: https://www.debates.org/voter-education/debate-transcripts/.

several interruptions and moments when the audience/interlocutors laugh throughout the debates, especially in the first one, between Donald Trump and Hillary Clinton. Consequently, laughter cannot be considered a marker of humorous intent here (see also Hay, 2001),[3] since the audience is not supposed to react to what is being said, which means that it is likely many humorous instances get no reaction from the public Nonetheless, it will be mentioned when the audience or the interlocutors laugh as a reaction to what is being said, because it marks an element of surprise and a clash between different layers of meaning.

The first debate contains 49 reactions that can be considered as including humorous intent, whereas the second debate includes 78 such reactions.[4] Interestingly, as will be shown in the Discussion part, the first debate includes more facial expressions used with these humorous exchanges than the second one (63.7% of the corpus), although the second debate includes more examples.

The humorous examples have been annotated using ELAN[5] regarding the facial/head markers used by the speakers when uttering them. Following the elements mentioned above, the attention falls on eyebrow movement (raised eyebrows and frowning), head movement (nods, shakes, tilts), shrugs and smiles. Hand movements were also used by the speakers, but they were not annotated in the corpus. When the face of the speaker/s was not visible during the humorous instances, it was marked as *not visible* in the corpus. This was mainly because of how the interlocutors were filmed during the debates, since the techniques are not consistent between the 2016 and the 2020 debate.

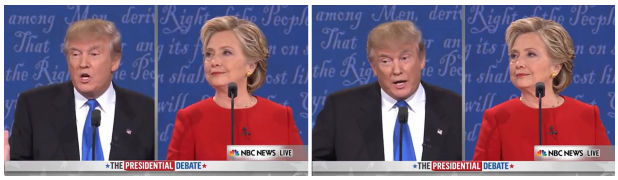
In the following section, examples from the corpus are presented to show how speakers make use of humor in order to build layers (Clark, 1996) of meaning and switch the viewpoint and the course of the conversation.

Examples: discourse markers

Several instances are presented here, with the use of non-verbal elements that have been discussed in relation to humor (see Tabacaru, 2019 or De Vries et al., 2024). These markers include raised eyebrows, nods, shrugs and smiles.

In the example below, Trump is being sarcastic toward some of Clinton's campaign strategies, such as the information available on her website, which she refers to for fact-checking purposes regarding his own claims. She refers to her website, which is followed by Trump's interruptions. The key elements are the words "her website" which allow a shift of focus from a meaning that was intended as [good] to something that is [bad] (similar to example [1] above where the negative evaluation is added afterwards). For the sake of relevance, some of the lines have been removed from the example. The humorous instance is marked in bold with the facial expressions in red below. The facial expressions in bold show that the non-verbal markers happen simultaneously:

(2) Clinton: So we have taken the home page of my website, HillaryClinton.com, and we've turned it into a fact-checker. […]
[...]
Trump: I think I should -- you go to her website, and you take a look at her website.
[...]
Trump: And look at her website. You know what? It's no different than this. She's telling us how to fight ISIS. Just go to her website. She tells you how to fight ISIS on her website.
I don't think General Douglas MacArthur would like that too much.
[raised eyebrows]
[nods]
[laughter from audience][1]



than this[2] too much

Facial expressions used in example (2)

[1]The video is available here: https://www.youtube.com/watch?v=qCHo67nqs4g (the example starts at 21m25s).

[2]("Libertarian Commentary during the First Debate between Donald Trump and Hillary Clinton", YouTube, uploaded by We Are Libertarians, 27 Sept 2016, https://www.youtube.com/watch?v=qCHo67nqs4g, Permissions: YouTube Terms of Service") (the example starts at 21m25s)

The sarcasm here is emphasized by Trump's critical words and behavior (facial expressions, nods) used to talk about the type of information provided by Clinton's website. The raised eyebrows used by Trump start when uttering *She tells you how to fight ISIS […]*,[6] which makes Clinton the target of his sarcasm (see also Tobin and Israel, 2012 or Tobin, 2020). He raises his eyebrows and nods repetitively when uttering "I do not think General Douglas MacArthur would like that too much." The mention of General Douglas MacArthur, military leader involved in both World War I and World War II, implies (Grice, 1975) that the strategies used by countries/leaders against their enemies should not be freely available online. The focus of attention is reversed: if, in Clinton's words, her website provides useful information regarding fact checking, Trump sarcastically refers to the other information that it provides, such as strategies regarding the war on terror. The reference to General Douglas MacArthur metonymically[7] gives access to the frame of successful military strategy, as opposed to Hillary Clinton's position.

Clinton also uses Trump's implied meaning in order to reverse the focus of attention once more, this time in her favor, but her facial expressions are different. Similar to findings by Gironzetti (2017) and Feyaerts et al. (2022), she uses a smile when implying that Trump does not have a plan regarding the war on terror, whereas she does:

3 De Vries et al. (2021) found that laughter happens more in ironic than non-ironic instances.

4 The identification of these instances was done using the theories and methods presented in Feyaerts (2013b) and Tabacaru, 2019. For example, Tabacaru (2019, p. 98) carried out reliability tests on humorous and non-humorous instances, with 84% agreement between the corpus and the participants to the reliability tests. It is safe to assume that humorous instances can be identified in either a written or oral corpus without the help of laughter.

5 ELAN is a tool for video annotation freely available for researchers from the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands (http://tla.mpi.nl/tools/tla-tools/elan/); see Brugman and Russel (2004), Wittenburg et al. (2006), and Sloetjes and Wittenburg (2008). I have used version 6.4.

6 The raised eyebrows are not visible at all times as the camera zooms out, but they are visible in the beginning and the end of his reaction. The screenshot here includes the very visible part at the end.

7 See, for example, Feyaerts (1999) or Tabacaru and Feyaerts (2016).

(3) Clinton: Well, at least I have a plan **to fight ISIS**.[1]
----------[smile][2]

Isis

Facial expression in example (3).

[1]Similar to the example above, the first part is not visible, as the camera is not filming the speaker, but the moderator.

[2]("Libertarian Commentary during the First Debate between Donald Trump and Hillary Clinton", YouTube, uploaded by We Are Libertarians, 27 Sept 2016, https://www.youtube.com/watch?v=qCHo67nqs4g, Permissions: YouTube Terms of Service") (the example starts at 22m33s)

This exchange continues, as both speakers turn the tables against each other using elements given by their opponent. In the following, Trump explicitly says that making the information available online is counterproductive because military strategies should not be accessed by the enemy, thus making explicit what he implicitly stated at the beginning of this exchange:

(4) Trump: No, no, you're telling the enemy everything you want to do. [Interruption]
Clinton: No, we're not. No, we're not.
Trump: See, you're telling the enemy everything you want to do. No wonder you've been fighting -- no **wonder** you've been fighting ISIS **your entire adult life**.
---------- [raised eyebrows] ---------- **[raised eyebrows]**
[nods]

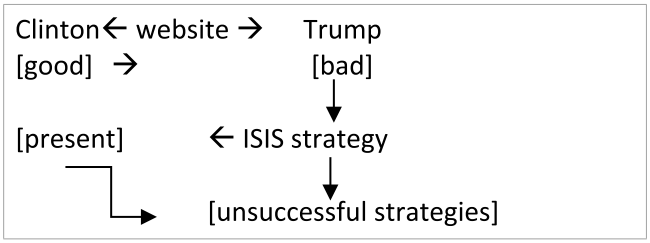
[laughter from audience][1]

your entire

Facial expression in example (4).

[1]("Libertarian Commentary during the First Debate between Donald Trump and Hillary Clinton", YouTube, uploaded by We Are Libertarians, 27 Sept 2016, https://www.youtube.com/watch?v=qCHo67nqs4g, Permissions: YouTube Terms of Service") (the example starts at 22m36s)

Trump here comes back to the information available on Clinton's website, further focusing on the fact that the enemy would have access to her military strategy. His last line emphasizes that Clinton is unsuccessfully fighting ISIS because she has made her plans public this entire time. The two speakers thus have different viewpoints regarding the war on terror, Clinton trying to focus on what is good about her website (the fact-checking option), while Trump emphasizes on what is bad, giving rise to different sarcastic instances from both speakers, as can be summarized below:



Both speakers here focus on what is beneficial to them, and it plays on different elements introduced by their interlocutor. The website is introduced by Clinton (see example [2] above) with a specific goal in mind, i.e., in order to propose a fact-checking option to verify the claims made, but Trump shifts the focus to the other information available there, such as military strategy, in order to show the negative side of the website. When Clinton picks up on this and defends her position by implying that, compared to him, she has already given this matter some thought, he further emphasizes the same idea, that she has been fighting the enemy unsuccessfully because they are already aware of her actions against them. Both interlocutors here enrich the common ground by defending their position and political actions (the case of Clinton) or by attacking their interlocutor (the case of Trump).

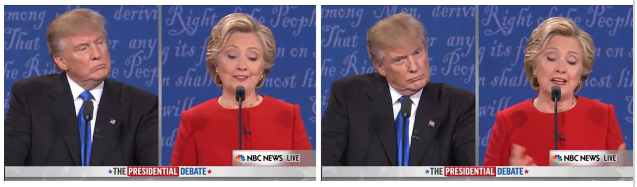
Raised eyebrows and nods are used in examples (2) and (4) with the sarcastic intentions of the speaker: when comparing Clinton's military strategy to that of General Douglas MacArthur, and when introducing the comment on [her] entire adult life, which is again a case of generalization (similar to example [1] above). The marker used by Clinton is different, as she uses a smile when aligning with what her opponent said, which also creates a pretense space in which Trump does not have any plan on the war on terror, whereas she does. The personal pronouns *she* and *I* are used by the two participants, the former acting as direct criticism and the latter adding a positive spin on oneself while indirectly criticizing the opponent (as a third-person reference).

Another similar example is (5) below, where comparisons between the two positions are presented. Here, Trump is remined by the moderator that he claimed Mrs. Clinton does not have a "presidential look," which he further argues using the word *stamina*. This word will be used by Clinton in her reply which emphasizes her experience in politics:

(5) Trump: She doesn't have the look. She doesn't have the stamina. I said she doesn't have the stamina, and I don't believe she does have the stamina. To be president of this country, you need tremendous stamina.
[...]
Clinton: **Well...**
----------[raised eyebrows] [shrug]
as soon as he travels to 112 countries
[raised eyebrows]
and negotiates a peace deal, a cease-fire, a release of dissidents, an opening of new opportunities in nations around the world **or even spends 11 hours** testifying in
----------[raised eyebrows] [----------
----------[shrug]
front of a congressional committee, he **can talk to me**
----------[shrug]

about **stamina**.
-------[nod]

[laughter and applause from audience][1]



Well... *or even*

Facial expressions for example (5).

[1]("Libertarian Commentary during the First Debate between Donald Trump and Hillary Clinton", YouTube, uploaded by We Are Libertarians, 27 Sept 2016, https://www.youtube.com/watch?v=qCHo67nqs4g, Permissions: YouTube Terms of Service") (the example starts at 1h29m59s)

When starting her answer to Trump's claim, Clinton uses a smile and raised eyebrows, as shown in above figure, which marks the start of her sarcastic response toward her opponent. This is also made manifest with the use of raised eyebrows again when uttering *or even* used to introduce other examples of her political experience. Other non-verbal elements include shrugs (after *Well*, with *or even* and *he can talk*), and nods when uttering the word *stamina*, which Trump had used.

Trump uses the word *stamina* four times to make a comment on his opponent's potential competence as future president of the United States,

to which Clinton answers using examples of her experience in politics, of which her opponent has none (the personal pronouns *she* and *he* are used by the two speakers to refer to each other directly). This, again, has to be interpreted against a shared common ground: Trump is not a politician, whereas she is. Her entire reply can be considered sarcastic as it represents a comparison to the absence of experience from Trump, but she uses certain discourse markers when she introduces her reply, in the middle of her argument to emphasize her claims, and again when uttering the word *stamina*, at the end.

Another example comes from the second debate, where, comparable to the repetition of the word *stamina* in the example above, Joe Biden repeats the word *segment*, used by the moderator, but adds a sarcastic meaning to it:

(6) Wallace: We have ended the segment. We're going to move on to the second segment.
Biden: That was really a productive **segment**, wasn't it? [to Trump] Keep yapping,
----------[nod]
man.
[moderator scoffs, laughter from audience after *wasn't it*][1]

[1]The video is available at https://www.youtube.com/watch?v=CweqW7Pzxz8; the example starts at 21m00s.

The word *segment* is uttered with a simple nod, which marks the repetition of the word used by the moderator, who had to intervene several times because of the interruptions between the speakers. Biden's implied meaning here is that they could not speak freely, without interruptions, which is further emphasized by his *Keep yapping, man*, aimed at his political adversary. The element of incongruity is very salient here, as Biden means the opposite of what he says: the segment that just finished was not productive as one of the speakers kept interrupting the other. Here, the participants and the audience have access to the same common ground that was created during the debate: the speakers kept talking at the same time, which means that their positions could not be successfully presented. Interestingly, the nod comes with the word *segment*, and not the adjective *productive*, which adds the pretense layer in Biden's remark. It is comparable to the example above where Clinton also nods when uttering stamina, the word previously used by her opponent. These represent key elements (Brône, 2008) that mark the switch between the two layers of meaning.

A final example with such repetitions is (7) below, in which the topic of taxes and jobs is discussed:

(7) Trump: But we have no leadership. And honestly, that starts with Secretary Clinton.
[Clinton laughs; audience laughter]
[...]
Clinton: I have a feeling that by, **the end of this evening,**
----------[smile]
I'm going to be blamed for **everything**[1]
----------**[raised eyebrows]**
----------**[long nod]**
that's ever happened.
[smile]
Trump: Why not?
----------[not visible]
[audience laughter]
Clinton: **Why not? Yeah, why not?**
----------[smile]
[audience laughter; Trump smiles][2]

[1]The word "everything" is also stressed by Clinton.

[2]The video is available here : https://www.youtube.com/watch?v=s7gDXtRS0jo from 24m22s.

Another case of generalization is the word *everything* sarcastically uttered by Clinton with marked non-verbal elements such as raised eyebrows, head movement, and a prosodic stress in order to accentuate that she is being accused of different things throughout the debate, such as is the case here. This follows Trump's position that he is not a politician, whereas she is (me versus them), not only a Democrat, but also Secretary of State, so, in a position of power and can metonymically refer to the frame of POWER. The key element *everything* marks sarcasm and exaggeration (Tabacaru, 2019), which is targeting her opponent, who keeps blaming her for different issues. This is very similar to examples (2) and (3) above, where Trump names her directly (through personal pronouns in the previous examples), whereas Clinton responds using the first person: *I have a feeling […] I'm going to be blamed*. If the focus of attention falls on her through her opponent's words, she shifts it through the use of exaggeration. The answer given by Trump, humorous as well, adds a layer not to the exaggeration to which she points, but to the word *blame*, with the rhetorical question meaning: *why should she not be blamed for everything*? Clinton then echoes Trump's words *why not* twice at the end, which are uttered with a smile, followed by a smile from Trump as well. All these lines represent sarcasm/humor, targeting the adversary, and all of them include non-verbal elements (except the one that is not visible, uttered by Trump).

Results

The results for the annotations regarding semiotic markers in the two debates considered here can be seen in Table 1. Unfortunately, in 13% of cases, the face of the speaker/s could not be seen (they were either not at all filmed or filmed from a distance, in which case the facial expressions could not be identified). There is also a clearer view in the first debate (filmed from the front) as opposed to the second debate, which included more examples of (humorous) interruptions, where the face of the speaker could not be seen. Similarly, the first debate between Trump and Clinton includes more non-verbal elements used with humor than the second one, although the number of humorous interactions was higher in the second debate compared to the first one.

For a clearer view of the discourse markers used in these humorous exchanges, the different eyebrows movements, head movements, and other markers have been considered together, as is shown in Figure 1. All eyebrow movements (raised eyebrows together, raising one eyebrow or frowning) and head movements (head tilts, shakes, or nods) are considered as two categories, for a general view of such resources.

These results show a preference for eyebrow and head movement, similar to other studies conducted on staged interactions (Tabacaru and Lemmens, 2014; Tabacaru, 2019), but also on spontaneous interactions (for example, De Vries et al., 2024). In 41.8% of cases of humor, speakers either used raised eyebrows or a frown, as shown in the examples presented above. The same goes for head movements, which are used in 34.8% of the data. The figure above also shows different behavior patterns for the two debates: both eyebrows and head movement are used more in the first debate (2016) than in the second one (2020), in which there were also more cases of markers that could not be analyzed

TABLE 1 Results of facial expressions in the two debates.

	Trump vs. Clinton	Trump vs. Biden	Total
Raised eyebrows	52	19	71
Nod/s	43	5	48
Not visible	3	25	28
Frown	7	8	15
Head shake/s	9	5	14
Head tilt	4	9	13
Shrug	7	4	11
Smile	10	1	11
Raised eyebrow	2	2	4
Total	137	78	215

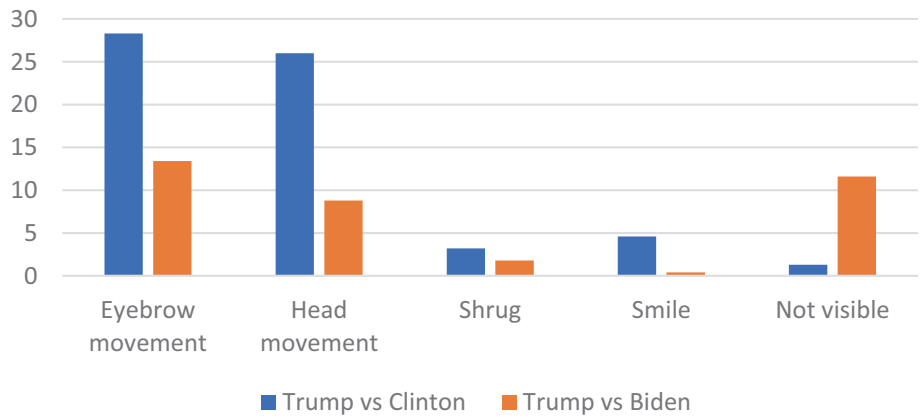


FIGURE 1
Results for discourse markers in the two debates.

for lack of visibility in the given videos (11.6% compared to 1.3%). Smiles are also used by the speaker/s, specifically in the first debate. This iscomparable to results presented by Gironzetti (2024), for example, although the intensity of such smiles has not been studied here, as has been the case elsewhere (see Gironzetti et al., 2016 or Ergül et al., 2024).

Of course, these markers are not specific to humor and are used in discourse in general (see also Tabacaru, 2019). For example, Jehoul et al. (2017) report that a shrug marks obviousness, which can also be the case here, but from a pretense space perspective. If we take example (5), Clinton shrugs several times during her answer to Trump's attacks: the first one can be considered a marker of obviousness, especially since it is used with "Well," but from a sarcastic perspective, i.e., it is obvious that Trump does not have her political experience. The shrug used with "he can talk to me [about stamina]" marks a pretense space, as it is implied that he will never have as much experience as she does. Debras (2013) mentions the shrug as marker of disengagement and this could also apply here, as Clinton seems to adopt an indifferent attitude toward Trump's attacks given her political experience (i.e., me versus him, as she also refers to him in the third person).

The examples presented here represent examples of sarcasm (which is considered a type of humor; see Tabacaru, 2019 or Tabacaru, 2020), where the target of the 'attack' is clear (Attardo, 2000, for example). The sarcastic meaning/implicature has then to be made clear with the use of both verbal and non-verbal markers.

Conclusion

This paper has explored the question of humorous stance in American political settings, of which two debates were presented in order to analyze how non-verbal markers were used by the speakers to make their meanings understood. Interactions such as these provide richer contexts for discourse analysis, although humor is not necessarily encouraged (this is the case for both debates analyzed here), as opposed to spontaneous interactions of humor between friends, for example. Nevertheless, the data shows the pervasiveness of humor, which can be used as a political tool in such contexts, as has been shown elsewhere (Li and Pryor, 2020, for instance). Humor is a mechanism that "appears frequently in interaction" (Priego-Valverde, 2024: 1), which seem to be case here as well. In a setting in which participants are not expected to 'amuse' the audience, humor still happens as a way to shift the

focus of attention (for example, from a [good] scenario to a [bad] scenario like is the case with the Covid-19 regulations) or to counter the adversary's attacks (as is the case with Clinton's focus on her political experience). As such, it is an interesting area of research in political settings.

The purpose of the paper was to show the way non-verbal discourse markers are used in interactions in order to create humorous meanings, and new layers built on the common ground (known or enriched during the debates). Such layers introduce pretense spaces of meaning aimed at the political opponent the speaker is debating. They also contribute to the topic of stance, as they show divergent viewpoints toward the same object/topic. In the case of political debates, humorous stance can thus play a role in the issue of persuasion because it emphasizes the interlocutors' ability to "turn the tables" on their opponent (Brône, 2008).

Although the videos used did not provide absolute view on the faces of the speakers, they contribute to showing the use of such markers with humorous intent. Similar to a key element in the case of hyper-understanding and misunderstanding (Brône, 2008), these non-verbal markers allow an emphasis on parts of speech that are relevant for a humorous interpretation. The results show a preference for both facial expressions and head movements, similar to staged interactions. Although the data is not balanced, neither in the way the speakers were filmed nor in the way they use these markers throughout the debates, they do provide a more detailed understanding on the way humor is a tool of human communication that is used for different purposes.

Data availability statement

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

Ethics statement

Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because videos are available online – copyright logo was added, and fair-use (US) and quotation rules (Europe) apply.

Author contributions

ST: Writing – original draft.

Funding

The author declares that financial support was received for the research, authorship, and/or publication of this article. The publication has been funded by UR TransCrit/Université Paris 8.

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

Andries, F., Meiss, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P. V., et al. (2023). Multimodal stance-taking in interaction—a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Attardo, S. (2000). Irony as relevant inappropriateness. *J. Pragmat.* 32, 793–826. doi: 10.1016/S0378-2166(99)00070-3

Attardo, S. (2003). Introduction: the pragmatics of humor. *J. Pragmat.* 35, 1287–1294. doi: 10.1016/S0378-2166(02)00178-9

Attardo, S., Pickering, L., and Baker, A. (2013). "Prosodic and multimodal markers of humor in conversation" in Prosody and humor. eds. S. Attardo, M. M. Wagner and E. Urios-Aparisi (Amsterdam: John Benjamins), 37–60.

Barnden, J. (2017). "Irony, pretence and fictively-elaborating hyperbole" in Irony in language use and communication. eds. A. Athanasiadou and H. Colston (Amsterdam/Philadelphia: John Benjamins Publishing), 145–177.

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press.

Bateman, J., Wildfeuer, J., and Hiippala, T. (2017). Multimodality: Foundations, research and analysis – A problem-oriented introduction. Berlin, Boston: De Gruyter.

Bertrand, R., and Priego-Valverde, B. (2011). Does prosody play a specific role in conversational humor? *Pragmatics Cogn.* 19, 333–356. doi: 10.1075/pc.19.2.08ber

Brône, G. (2008). Hyper and misunderstanding in interactional humor. *J. Pragmat.* 40, 2027–2061. doi: 10.1016/j.pragma.2008.04.011

Brône, G., and Oben, B. (2013). "Resonating humor: a corpus-based approach to creative parallelism in discourse" in Creativity and the agile mind: A multidisciplinary approach to a multifaceted phenomenon. eds. T. Veale, K. Feyaerts and C. Forceville (Berlin: Mouton de Gruyter), 181–203.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., and Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cogn. Linguist.* 28, 449–483. doi: 10.1515/cog-2016-0119

Brugman, H., and Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation.

Clark, H. H. (1996). Using language. Cambridge: Cambridge University Press.

Clark, H. H., and Gerrig, R. J. (1984). On the pretense theory of irony. *J. Exp. Psychol.* 113, 121–126. doi: 10.1037/0096-3445.113.1.121

Cobb, M. D., and Kuklinski, J. H. (1997). Changing minds: political arguments and political persuasion. *Am. J. Polit. Sci.* 41, 88–121. doi: 10.2307/2111710

Dancygier, B., and Vandelanotte, L. (2016). "Discourse viewpoint as network" in Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities. eds. B. Dancygier, L. Wei-lun and A. Verhagen (Berlin: De Gruyter Mouton), 13–40.

De Vries, C. M., Clarissa, M., Oben, B., and Brône, G. (2021). Exploring the role of the body in communicating ironic stance. *Lang. Modal.* 1, 65–80. doi: 10.3897/lamo.1.68876

De Vries, C. M., Clarissa, M., Oben, B., and Brône, G. (2024). "On target. On the role of eye-gaze during teases in face-to face multiparty interaction" in Interactional humor: Multimodal design and negotiation. ed. B. Priego-Valverde (Berlin: Mouton de Gruyter), 53–85.

Debras, C. (2013). Multimodal stance-taking in a videotaped corpus of discussions about environmental issues in British English. PhD dissertation, defended on 07/12/2013. Paris: University Paris 3.

Du Bois, J. (2007). "The stance triangle" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam: John Benjamins), 139–182.

Ekman, P., and Wallace, F. V. (2003). Unmasking the face. A guide to recognizing emotions from facial expressions. Cambridge, MA: Malor Books.

Englebretson, R. (2007). "Grammatical resources for social purposes some aspects of stancetaking in colloquial Indonesian conversation" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam: John Benjamins Publishing Company), 69–110.

Ergül, H., Miller, S., Attardo, S., and Kramer, K. (2024). "Alternative conceptualizations of the smiling intensity scale (SIS) and their applications to the identification of humor" in Interactional humor: Multimodal design and negotiation. ed. I. B. Priego-Valverde (Berlin: Mouton de Gruyter), 109–130.

Feldman, O. (2024). "Humor and politics: a conceptual introduction" in Political humor worldwide. The language of politics. ed. O. Feldman (Singapore: Springer).

Feyaerts, K. (1999). "Metonymic hierarchies: the conceptualization of stupidity in German idiomatic expressions" in Metonymy in language and thought. eds. K.-U. Panther and G. Radden (Amsterdam/Philadelphia: John Benjamins Publishing), 309–332.

Feyaerts, K. (2013a) in A cognitive grammar of creativity. Creativity and the agile mind: A multidisciplinary approach to a multifaceted phenomenon. eds. T. Veale, K. Feyaerts and C. Forceville (Berlin: Mouton de Gruyter), 205–227.

Feyaerts, K. (2013b). "Tackling the complexity of spontaneous humorous interaction. An integrated classroom-modeled corpus approach" in Irony and humor. eds. L. Ruiz-Gurillo and M. B. A. Ortega (Amsterdam & Philadelphia: John Benjamins Publishing), 243–268.

Feyaerts, K., Brône, G., and De Ceukelaire, R. (2015). "The art of teasing. A corpus study of teasing sequences in American sitcoms between 1990 and 1999" in Cognitive linguistics and humor research. eds. G. Brône, K. Feyaerts and T. Veale (Berlin: Mouton de Gruyter), 215–243.

Feyaerts, K., Brône, G., and Oben, B. (2017). "Multimodality in interaction" in The Cambridge handbook of cognitive linguistics. ed. B. Dancygier (Cambridge: Cambridge University Press), 135–156.

Feyaerts, K., and Oben, B. (2014). "Tracing down schadenfreude in spontaneous interaction. Evidence from corpus linguistics" in Schadenfreude': Understanding pleasure at the misfortune of others. eds. W. Van Dijk and J. W. Ouwerkerk (Cambridge: Cambridge University Press), 275–291.

Feyaerts, K., Rominger, C., Lackner, H. K., Brône, G., Jehoul, A., Oben, B., et al. (2022). In your face? Exploring multimodal response patterns involving facial responses to verbal and gestural stance-taking expressions. *J. Pragmat.* 190, 6–17. doi: 10.1016/j.pragma.2022.01.002

Geeraerts, D. (2021). Second-order empathy, pragmatic ambiguity, and irony. In SilvaA. S. da (Ed.), Figurative language–Intersubjectivity and usage: 19–40. Amsterdam: John Benjamins Publishing Company.

Gironzetti, E. (2017). "Prosodic and multimodal markers of humor" in The Routledge handbook of language and humor. ed. S. Attardo (London: Routledge), 400–413.

Gironzetti, E. (2022). The multimodal performance of conversational humor. Series figurative language and thought 13. Amsterdam: John Benjamins Publishing Company.

Gironzetti, E. (2024). "Humorous smiling: a reverse cross-validation of the smiling intensity scale for the identification of conversational humor" in Interactional humor: Multimodal design and negotiation. ed. B. Priego-Valverde (Berlin & Boston: Mouton de Gruyter), 87–107.

Gironzetti, E., Attardo, S., and Pickering, L. (2016). "Smiling, gaze, and humor in conversation" in Metapragmatics of humor: Current research trends. ed. L. Ruiz-Gurillo (Amsterdam: John Benjamins Publishing Company), 235–254.

Grice, P. (1975). "Logic and conversation," in Figurative language–Intersubjectivity and usage. eds. P. Cole and J. L. Morgan (New York: Academic Press) 3, 41–58.

Hay, J. (2000). Functions of humor in the conversations of men and women. *J. Pragmat.* 32, 709–742. doi: 10.1016/S0378-2166(99)00069-7

Hay, J. (2001). The pragmatics of humor support. *Humor* 14, 55–82. doi: 10.1515/humr.14.1.55

Hunston, S. (2007). "Using a corpus to investigate quantitatively and qualitatively" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam & Philadelphia: John Benjamins Publishing Company), 27–48.

Jehoul, A., Brône, G., and Feyaerts, K. (2017). The shrug as marker of obviousness: corpus evidence from Dutch face-to-face conversations. *Linguist. Vang.* 3, 1–9. doi: 10.1515/lingvan-2016-0082

Langacker, R. W. (1987). Foundations of cognitive grammar, i. theoretical prerequisites. Stanford, California: Stanford University Press.

Li, X. (2021). Multimodal practices for negative assessments as delicate matters: incomplete syntax, facial expressions, and head movements. *Open Linguistics* 7, 549–568. doi: 10.1515/opli-2020-0164

Li, S., and Pryor, T. (2020). Humor and persuasion: the effects of laughter during US supreme Court's Oral arguments. *Law Policy* 42, 162–185. doi: 10.1111/lapo.12145

Mondada, L. (2009). The embodied and negotiated production of assessments in instructed actions. *Res. Lang. Soc. Interact.* 42, 329–361. doi: 10.1080/08351810903296473

Norris, S. (2004). Analyzing multimodal interaction: A methodological framework. New York: Routledge.

Priego-Valverde, B. (2009). "Failed humor in conversation: a double voicing analysis" in Humor in interaction. eds. N. R. Norrick and D. Chiaro (Amsterdam: John Benjamins Publishing), 166–183.

Priego-Valverde, B. (2024). Interactional humor: Multimodal design and negotiation. Berlin: Mouton de Gruyter.

Raskin, V. (1985). Semantic mechanisms of humor. D. Dordrecht: Reidel.

Scheibman, J. (2007). "Subjective and intersubjective uses of generalizations in English conversations" in Stancetaking in discourse: Subjectivity, evaluation, interaction. ed. R. Englebretson (Amsterdam & Philadelphia: John Benjamins Publishing Company), 111–138.

Sloetjes, H., and Wittenburg, P. (2008). Annotation by category—ELAN and ISO DCR. Proceedings of the 6th international conference on language resources and evaluation, LREC.

Stubbs, M. (1986). "A matter of prolonged fieldwork: Notes towards a modal grammar of English," *Applied Linguistics*. 7, 1–25.

Tabacaru, S. (2019). "A multimodal study of sarcasm in interaction" in Series applications of cognitive linguistics 40 (Berlin: Mouton de Gruyter).

Tabacaru, S. (2020). "Faces of sarcasm. Exploring raised eyebrows with sarcasm in French political debates" in Diversity of irony. eds. D. A. Athanasiadou and H. L. Colston, 256–277.

Tabacaru, S., and Feyaerts, K. (2016). The power of metonymy in humor: stretching contiguous relations across different layers of meaning. *European J. Humor Res.* 4, 1–18. doi: 10.7592/EJHR2016.4.2.TABACARU

Tabacaru, S., and Lemmens, M. (2014). Raised eyebrows as gestural triggers in humor: the case of sarcasm and hyper-understanding. *European J. Humor Res.* 2, 11–31. doi: 10.7592/EJHR2014.2.2.tabacaru

Tobin, V. (2020). "Experimental investigations of irony as a viewpoint phenomenon" in The diversity of irony. eds. A. Athanasiadou and H. L. Colston (Berlin: De Gruyter Mouton).

Tobin, V., and Israel, M. (2012). "Irony as a viewpoint phenomenon" in Viewpoint in language: A multimodal perspective. eds. B. Dancygier and E. Sweetser (New York: Cambridge University Press), 25–46.

Tomasello, M. (1999). The cultural origins of human cognition. Cambridge: Harvard University Press.

Verhagen, A. (2005). Constructions of intersubjectivity: Discourse, syntax, and cognition. Oxford: Oxford University Press.

Wittenburg, P., Brugman, H., Russell, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.

Yus, F. (2003). Humor and the search for relevance. *J. Pragmat.* 35, 1295–1331. doi: 10.1016/S0378-2166(02)00179-0

Zhang, Q. (1998). Fuzziness – vagueness – generality – ambiguity. *J. Pragmat.* 29, 13–31. doi: 10.1016/S0378-2166(97)00014-3

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 April 2025
DOI 10.3389/fpsyg.2025.1509988

Check for updates

OPEN ACCESS

EDITED BY
Elisabeth Zima,
University of Freiburg, Germany

REVIEWED BY
Steven Schoonjans,
University of Klagenfurt, Austria
Susanne Reinhardt,
University of Potsdam, Germany

*CORRESPONDENCE
Emma Betz
✉ embetz@uwaterloo.ca
Alexandra Gubina
✉ gubina@ids-mannheim.de

†These authors have contributed equally to this work and share first authorship

RECEIVED 11 October 2024
ACCEPTED 10 February 2025
PUBLISHED 28 April 2025

CITATION
Betz E and Gubina A (2025) On stance-taking with one-sided vs. two-sided shoulder lifts in German talk-in-interaction.
Front. Psychol. 16:1509988.
doi: 10.3389/fpsyg.2025.1509988

COPYRIGHT
© 2025 Betz and Gubina. 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.

On stance-taking with one-sided vs. two-sided shoulder lifts in German talk-in-interaction

Emma Betz[1*†] and Alexandra Gubina[2*†]

[1]Department of Germanic and Slavic Studies, University of Waterloo, Waterloo, ON, Canada, [2]Leibniz Institute for the German Language, Mannheim, Germany

Taking a stance toward events, objects, and other persons is fundamental to human interaction. We investigate one specific body movement that is involved in stance-taking in interaction: a shoulder lift, realized as either a one-sided or a two-sided movement. Using multimodal Conversation Analysis, we trace how interactants employ shoulder lifts in different positions within responsive turns in various interaction types in German. This study reveals how the actions to which shoulder lifts contribute are bound to specific turn and sequence positions. We demonstrate how shoulder lifts are used for disclaiming the speaker's accountability or responsibility by framing their turn as non-expandable or non-expansion-worthy, thus curtailing the sequence. Furthermore, the study shows how participants orient to different types of shoulder movements, i.e., lifts with one or with both shoulders, as accomplishing different interactional tasks. By showing that shoulder lifts are a positionally sensitive resource for speakers in building stances, we showcase the potential of conversation analytic and interactional linguistic approaches to further our understanding of multimodal stance-taking in interaction.

KEYWORDS

shrug, shoulder lift, embodiment, stance, accountability, responsibility, resistance, multimodal Conversation Analysis

1 Introduction

Taking a stance toward events, objects, and other persons is fundamental to human interaction. Stance-taking can foreground matters of knowledge, rights/responsibilities to act, and affect (e.g., Du Bois, 2007; Heritage, 2012; Goodwin and Alim, 2010; Stevanovic and Peräkylä, 2014). Stance-taking involves resources of language and body. Prior interactional research has concentrated on linguistic practices for stance-taking; work on embodied practices has started to emerge recently (e.g., Cekaite, 2010; Clift, 2024; Ford et al., 2012; Heller et al., 2023; Kaukomaa et al., 2014; Marrese et al., 2021). Our paper contributes to this latest line of research by investigating a specific body movement that is involved in stance-taking in interaction—a shoulder lift, realized as either a one-sided or a two-sided movement. Using multimodal Conversation Analysis (CA; Sidnell et al., 2012; Mondada, 2018), we trace how interactants employ shoulder lifts in different positions within responsive turns in a range of interaction types in German.

Prior research has described shoulder lifts (commonly understood to involve both shoulders) as one core component of "shrugs" (Debras, 2017; Givens, 1977; Morris, 1994; Streeck, 2009)—complex ensembles including such elements as head tilts, shoulder lifts, and certain mouth configurations. It has been suggested that the "shrug" has a more general unified meaning (*disengagement*, Streeck, 2009, pp. 189–91; cf. Debras and Cienki, 2012), and that this can convey distinctly different stances: incapacity and non-responsibility, affective distance or indifference, epistemic meanings like indetermination and common ground (Debras, 2017).

The present study heeds the call for more research on the body movements commonly associated with "shrugs" in different languages and in a broader range of data, specifically in spontaneous, unguided, naturally occurring interactions. We analyze shoulder lifts in their precise sequential environments and ask (a) where/how speakers use different types of shoulder lifts (i.e., one-sided vs. two-sided) systematically in real-time interaction, (b) what actions lifts contribute to and what stances they (contribute to) convey(ing), and (c) whether there are differences in use between different types of lifts in different positions within responsive turns.

The following data extract includes three separate shoulder lifts and offers a glimpse of the range of our phenomenon. It comes from a mealtime interaction between friends Gero, Zoe (seated to Gero's left), and Zoe's boyfriend Norbert (seated across from Gero). In lines 01–02, Gero initiates a new topic and sequence by announcing that a new intern has joined their department. In line 3, Norbert first acknowledges the news and then pursues assessment by Gero with a rising-intoned *und* ("and"). In response, Gero first produces the vocalization *pf* (line 5; see Baldauf-Quilliatre and Imo, 2020) that projects a negative assessment, which is uttered in line 8. Norbert responds by laughing, and Zoe criticizes Gero for always judging "them" (meaning either female interns or women in general) based on their outward appearance (lines 10–11).

In response, Gero produces three different types of shoulder lifts:

- A two-sided movement accompanying his account in line 13 (Figures 1–3),
- a two-sided movement after catching Zoe's gaze in line 15 (Figures 4–6),
- and a one-sided movement with his concession in line 16 (Figures 7–9).

Extract (1): FOLK_E_00293_SE_01_T_04_DF_01_c913_zweitrangig

http://bit.ly/3BS40hu

```
01     GER:  ((smacks)) wir ham jetz auch ne prak- ne (.)
                        we now also have an int- a

02           brasilianische praktikantin.
             Brazilian intern

03     NOR:  ja, (.) und?
             yes     and

04           (0.3)

05     GER:  pf:_

06           (0.83)

07     NOR:  nh nh nh nh nh ((laugh through nose))

08     GER:  voll hässlisch.
             really ugly

09     NOR:  ((laughs for 1.7 sec.))

10     ZOE:  dass du die immer (alle)
             that you always

11           nach dei nach der ↑optik↑ ↓bewerten musst.
             have to assess them (all) by you- by their looks

12           (0.5)#
       fig        #1
13 => GER:  *ja #ihr fAchlisch kompetenz kenn isch *net*.#
             (well) her/their professional competence I'm not familiar with
       ger   *lifts both shoulders-----------------*,,,*
       fig       #2                                  #3
```

Figure 1

Figure 2

Figure 3

14 NOR: °nh nh nh°

15 => #(2.0)#*(1.5) *#(2.9)#
ger *lifts both shoulders*
fig #4 #5 #6 #7

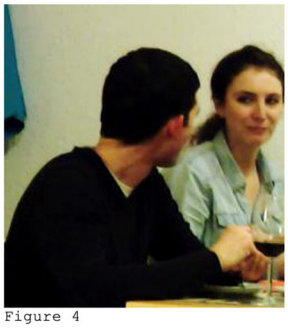


Figure 4

Figure 5

Figure 6

16 => GER: ***#dies is* mir #aber auch relativ egal,**
but it doesn't really matter to me anyway
ger *lifts left shoulder*
fig #8 #9

Figure 7

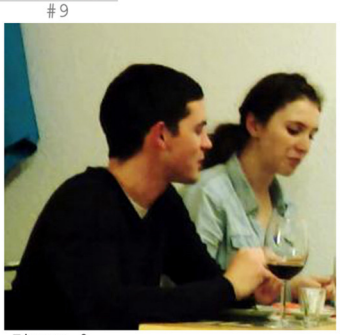


Figure 8

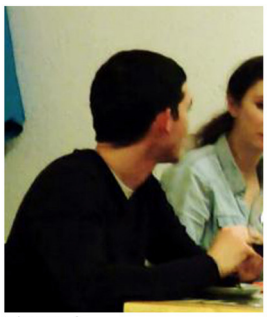


Figure 9

17 (0.3)

18 ZOE: hm hm, (.) wollt_s grad sagen. (.) is eher zweitrangig.=ne,
hm_hm was just gonna say that it's rather secondary, right?

19 (1.3)

Example 1 shows that shoulder lifts occur in different turn positions: in responsive position accompanying talk, before a possible turn completion, in "pre-beginning" position, and in "post-possible completion" position (Schegloff, 1996a, pp. 90–93). In some cases, shoulder lifts can be produced not only within responsive turns but as *responsive turns per se*. In this paper, we show that shoulder lifts in these different positions accomplish various kinds of interactional functions, but we also reveal that they share commonalities: In all these cases, shoulder lifts are used for

- disclaiming the speaker's accountability, or responsibility (in the case of Extract 1 for their own behavior) through
- framing their turn as non-expandable, or non-expansion-worthy, and thus
- curtailing the sequence.

Furthermore, we will show that the interactional work of a shoulder lift and thus its meaning depend not only on where it occurs within a TCU/turn (position), but also how it is produced (composition), i.e., as a one-sided or two-sided lift. Specifically, this study will demonstrate that movements with one or with both shoulders are employed for—and afford for—accomplishing different interactional work across various settings and sequential positions: While two-sided shoulder lifts are systematically employed for indexing the speakers' lack of *ability* to "go along," or to further engage with the Other's initiated or projectable course of action (e.g., due to lack of knowledge), one-sided shoulder lifts tend to mark "non-engagement," in other words, *resistance* to "go along" with the Other's initiated or projectable course of action.

2 Paralinguistic and bodily resources for stance-taking in talk-in-interaction

Taking a stance—be it toward a material object, a person, or a turn-at-talk—is omnirelevant in talk-in-interaction, i.e., "there is never a time out from the social action of taking stances and adopting positions" (Du Bois and Kärkkäinen, 2012, p. 438). By taking a stance, participants publicly position themselves toward a

stance object through evaluating it. In doing so, they either align or misalign with co-interactants' stances (Du Bois, 2007; Du Bois and Kärkkäinen, 2012). Most studies within Conversation Analysis and Interactional Linguistics on stance-taking (see Stevanovic and Peräkylä, 2014) concentrate on *linguistic* means for stance-taking. However, other resources for taking a stance in talk-in-interaction have received increasing attention across different languages (see Andries et al., 2023 for a comprehensive overview).[1]

Stance-taking can be accomplished with paralinguistic means, which include non-lexical vocalizations, or sound objects, i.e., "conversational objects with [phonetic substance but] minimal semantic content." These include "so-called 'primary interjections,' such as, e.g., *oh, ah* and *ooh*, and non-lexical sounds such as clicks and whistling, which have been found to function similarly in talk-in-interaction" (Reber, 2012, p. 12; Keevallik and Ogden, 2020). Some of these "sound objects" are strongly associated with displaying certain emotions and accomplishing affective work, such as expressing surprise (Wilkinson and Kitzinger, 2006) or disappointment (Couper-Kuhlen, 2009), expressing "suffering" with moaning, whining or crying (Edwards, 2005; Hepburn and Potter, 2007), conveying a sense of pleasure with gustatory *mmm* (Wiggins, 2002; Golato, 2011), or conveying disgust with *eugh* (Wiggins, 2013). Interactants can also position a specific referent as laughable with laughter (Jefferson et al., 1977; Glenn and Holt, 2013) or convey a negative stance toward an assessment with the "sound object" *pf* (see Extract 1, line 5), thus indicating that they lack the words to verbally express the assessment in a different way (Baldauf-Quilliatre and Imo, 2020, p. 212).

Another group of non-lexical vocalizations that can display affective stance are respiratory phenomena like whistles, sighs, and clicks. In their study on melodic and non-melodic whistles, Reber and Couper-Kuhlen (2020) demonstrate that a non-melodic gliding whistle that "begins on a high tone and glides slowly downwards" (p. 177) systematically occurs in response to informings that describe a norm-breaching state of affairs and often contains a numerical reference. The gliding whistle conveys a specific affective stance toward the prior informing: It treats it as impressive and/or shocking (see also Reber, 2012, ch. 9.2). Clicks constitute another interactional resource that is used for stance management in talk-in-interaction (Reber, 2012, ch. 9.1; Wright, 2011; Ogden, 2013, 2020). Reber demonstrates that clicks produced in response to complaints can index "disapproval" of the complainable and thus display "a concordant and hence affiliative stance, yet without bringing in too much affective involvement" (Reber, 2012, p. 234). Ogden (2020) shows how clicks are used as a practice for audibly not accomplishing socially improper actions (like self-praise or sexual allusions) that could "put the (non-)speaker in an awkward or conflictual position" (Ogden, 2020, p. 86). Sniffing can project dispreferred responses or be part of "delicate" actions (Hoey, 2020). Sighing is routinely and indexically associated with negative affect and can thus project a negative valence (Hoey, 2014).

Taking or projecting a stance can also be accomplished through bodily means. These include facial expressions (Groß et al., 2023). Turn-beginning frowns in particular have been shown to anticipate turns that are potentially problematic and can constitute a complication, like an upcoming negative assessment, disaffiliation, or epistemic difficulty (Kaukomaa et al., 2014). Turn-opening smiles can project a "shift from a neutral or serious emotional stance to a positive or humorous emotional stance" (Kaukomaa et al., 2013, p. 21). And an eyeroll can display protest or dissent with the prior speakers' actions, as Clift (2021) demonstrates.

Recent studies have also focused on how multimodal packages - combinations of verbal, vocal, and bodily resources - can display specific stances. Clift (2023) demonstrates how "raising of the eyebrows—which has the effect of furrowing the brow—in conjunction with a pursing of the lips" (Clift, 2023, p. 172) can be treated as a display of skepticism. Similar results were obtained by Heller et al. (2023), who show how bodily, verbal, and prosodic resources are employed in tandem to display a critical stance in two actions: contradicting another and calling something into question. Contradicting is characterized by contracting and lowering eyebrows and by narrowing of the eyes, gaze aversion from the recipient, and nose-wrinkling. Questioning, by contrast, consistently occurs with prolonged contractions and an upward or downward movement of the eyebrows, as well as direct gaze at recipients. Depending on whether the eyes are narrowed or wide and depending on which specific prosodic resources are employed, the stance displayed through questioning can be more confrontational or less so.

Stance displays can also be accomplished with embodied resources other than facial expressions. Clift (2014) describes a practice called "visible deflation," which is characterized by "a bodily 'let-down' from a position in tension" (Clift, 2014, p. 381) and used for displaying a negative stance—"exasperation"—in response to others' actions. In a study on complaints in French, Skogmyr Marian (2021) demonstrates that specific types of embodied conduct can be used to complete incomplete verbal turns and display a negative affective stance. Such resources encompass depictive gestures (Streeck, 2009) that have a conventional negative meaning, embodied gestures associated with expressing negation and negative stance (Kendon, 2002), or pragmatic gestures (Kendon, 2004) such as the placing of hands on the table. Our article continues this line of research by focusing on shoulder lifts and their role in stance-taking in social interaction.

3 Shoulder lifts in interaction

As noted in Section 1, prior research has treated shoulder lifts as the most prototypical, core component of shrugs (Streeck, 2009; Debras and Cienki, 2012), which is why "shoulder lift" and "shrug" are often used as synonymous descriptions. Earlier studies on shrugs have classified them as "emblems" (Ekman and

1 It should be noted that such studies primarily concentrate on affective, or emotional, stance (but see, e.g., Cekaite, 2010; Deppermann and Gubina, 2021; Goodwin and Cekaite, 2013; Heller, 2018; Schoonjans, 2018 on bodily and paralinguistic means for displaying epistemic and deontic stance).

Friesen, 1969) or "quotable gestures" (Kendon, 2004, p. 335), i.e., as gestures with a fixed, conventionalized meaning within a specific culture. In contrast, recent, more interactionally-grounded approaches have treated shrug as a complex enactment, in which "a single component or a combination of components can index the whole enactment" (Debras, 2017, p. 1; Streeck, 2009, p. 190). Possible components include raised eyebrows, palm-up gestures, head tilts, mouth shrugs (i.e., when the corners of the mouth turn downward and the middle part of the mouth is raised, see Extract 5, Figure 14 for an example), possibly in combination with an upward-forward chin movement and shoulder lifts. However, it is important to note that single components can vary in the degree of specificity of their meaning, e.g., a mouth shrug has been associated with the specific meaning of epistemic negation or ignorance (Givens, 1977, p. 13; Morris, 1994, p. 65; Streeck, 2009, p. 190), while shoulder lifts can accomplish other functions, as we will show in Section 5.

Prior research has also proposed a general, or context-free, meaning of shrugs, namely *disengagement* (Kendon, 2004, p. 265; Streeck, 2009, p. 190; Debras, 2017) or "*dis-stance*": Speakers position themselves "with respect to a prior stance, while simultaneously acknowledging [a] stance differential" (Debras and Cienki, 2012, p. 6; see Du Bois, 2007) between their positioning and that of other interlocutors (see also Schegloff, 1996a, p. 92, who lists shrugs as stance markers that can appear at possible turn completion). This general meaning has also been specified as displaying a participant's ignorance or uncertainty (Tennant and Brown, 1998, p. 180), non-intervention to whatever is in focus (Kendon, 2004, p. 275), or non-assertiveness (Givens, 1977), and as functioning as a disclaimer (Morris, 1994 on facial shrugs).

More interaction-oriented approaches (Debras, 2017, p. 12) differentiate between three semantic categories, or meanings, of shrugs: (1) incapacity and non-responsibility, (2) affective distance or indifference, (3) epistemic meanings like expressing indetermination and indexing common ground. Shrugs have also been shown to accomplish functions similar to epistemic-evidential markers, "which relate[...] to the gesturer's degree of knowledge of, and commitment to, a state of affairs, and to the origin of a gesturer's knowledge" (Debras and Cienki, 2012, p. 936). For instance, shoulder shrugs (together with direct gaze on recipients) co-occur with formats like *je (ne) sais pas* (Debras, 2021; Pekarek Doehler, 2022), indexing lack of epistemic access in response to questions. Further functions in question are uncertainty, obviousness (see Jehoul et al., 2017; Schoonjans, 2018, pp. 160–64), and logical or chronological necessity.

As this overview demonstrates, prior research does not systematically differentiate between shrugs as complex configurations and shoulder lifts as one particular resource with possibly distinct features (e.g., one-sidedness) used for accomplishing specific interactional work. Debras (2017, p. 17) acknowledges a possible perceptual difference between one-sided and two-sided shoulder lifts by describing one-sided shrugs as "less conspicuous," but it remains unclear whether participants actually orient differently to one- and two-sided shoulder shrugs (i.e., as accomplishing different interactional tasks). Moreover, interactionally oriented research on shoulder shrugs is based primarily on semi-guided conversations (Debras, 2017; Debras and Cienki, 2012), which is why we still know little about how and when shoulder lifts appear in naturally-occurring interaction. The present study addresses these open questions by systematically investigating the use of one-sided and two-sided shoulder lifts in naturally occurring, unelicited and unguided interactions in German. We seek to determine how different types of shoulder lifts are employed by participants in different positions within turns and sequences to accomplish, or contribute to, distinct interactional functions. Moreover, we will demonstrate that shoulder lifts are different from most of the para- and non-verbal resources used for stance-taking already described (see Section 2) in terms of their *stance object*: In contrast to other non-verbal resources that can be used for taking a stance, shoulder lifts allow speakers to position themselves not toward *what another participant said or did,* but rather toward *their own utterance(s)*. By analyzing a diverse set of interaction data, we explore the nuanced roles that shoulder lifts play in stance-taking and in talk-in-interaction in general.

4 Data and methods

Research on multimodal interaction has shown that the precise temporal ordering of embodied behavior and talk is constitutive for social action. Using multimodal CA (Mondada, 2018), we trace how interactants employ shoulder lifts in different turn positions in a range of interaction types in German. Our data come from the German FOLK corpus[2] and are drawn from more than 80 h of video recordings of naturally occurring face-to-face interaction in everyday and institutional contexts (e.g., driving lessons, physiotherapy sessions) in both stationary settings (e.g., family breakfasts, playing table-top games) and mobile configurations (e.g., joint activities such as cooking or renovating a room). A small number of additional examples were added as we came across them while working on other projects; these are from private corpora (including mundane and workplace interaction, face-to-face and video-mediated) and from public interaction (e.g., political debates). All non-public data were collected with participants' informed consent, and all names appearing in the transcripts have been anonymized.

The initial collection we assembled consisted of 259 cases, of which 108 were identifiable as two-sided shoulder lifts and 118 were one-sided lifts. In 33 cases, we could—due to the camera angle, the quality of the recording, or the spatial arrangement of participants—not determine with confidence whether one or two shoulders were lifted, or whether we were dealing with a communicative shoulder lift at all (as opposed to a shoulder movement that resulted from adjusting a different part of the body or from the body's involvement in laughter). These examples were set aside. CA methodology (Schegloff, 2007; Sidnell et al.,

2 The corpus is hosted at the *Leibniz Institute for the German Language* and is accessible to scholars after registering at http://dgd.ids-mannheim.de. For a detailed description of the German FOLK-corpus, see Schmidt (2016).

2012; Robinson et al., 2024) allowed us to relate shoulder lifts to the precise turn and sequential positions in which they occur in the back and forth of real-time interaction. Since we were interested in participants' orientations to the actions and positionings accomplished by turns that consisted of/contained shoulder lifts (on participant orientations, see Raymond and Robinson, 2024), we limited our focus to broadly responsive uses. These include shoulder lifts that stand alone *as* responses, those that accompany responsive verbal turns, and those that are produced at the boundaries of responsive turns (i.e., shoulder lifts that precede verbal turns, complete them, or follow possibly complete verbal turns). We did not limit our focus to specific action environments in which shoulder lifts occur. However, most of the shoulder lifts in our collection are produced in response to informings; requests for information or confirmation; as well as challenges, reproaches, and account solicitations. Due to space limitations, we excluded shoulder lifts produced as part of multi-unit turns (e.g., story-tellings, 139 instances).[3] The resulting collection of target cases for this study contained 87 instances (44 one-sided and 43 two-sided shoulder lifts).

The examples assembled in this collection were examined individually in detail—including the shape of the shoulder lifts and their precise placement in turn, sequence, and larger activity—and compared to each other to uncover commonalities in formal, sequential, and action patterns (on working with collections, see Clayman, 2024). In an iterative process, we were able to identify distinct contexts of use and stable ways in which shoulder lifts contribute to stance-taking in interaction. We analyzed around 50 cases in this way before reaching a point of analytic saturation (Corbin and Strauss, 1990). We then checked each of the remaining examples in the collection to test and refine our findings and to determine the breadth and variability of the practice.

In Section 5, we report our findings and show representative examples of the different uses of shoulder lifts we uncovered in spontaneous interaction. The extracts were transcribed according to Jeffersonian and Mondada transcription conventions (Jefferson, 2004; Mondada, 2024), and talk is presented in the original German with idiomatic translation into English. The shoulder lifts are highlighted in gray and shown in green font. Responses that are accompanied, preceded, or followed by shoulder lifts are marked with an arrow (=>) on the left. Each extract is accompanied by a link to the video data (see Footnote 2 for corpus access details). In inspecting the data, readers will notice that shoulder movements are quite prominent in some instances (e.g., Extracts 11 *dahergesetzt*, 13 *nervich*) and small and initially hard to see in others (e.g., Extract 2 *sixt*). Furthermore, some movements we classified as one-sided may not initially appear to be one-sided. This is because in all shoulder movements, other parts of the upper body are invariable involved, so that one can observe slight movement of the other shoulder in many one-sided shoulder lifts. In analyzing our examples, we used differences in the prominence of movement to determine whether we are dealing with two- or one-sided lifts. A one-sided shoulder lift in our collection is one in which the movement of (only) one side seems deliberate and is more prominent.

3 The beginning of Extract (4) shows such shoulder lifts. They are produced by the the driving instructor (see line 3).

5 Shoulder lifts in different turn positions in interaction

This section is organized to represent different positions in which shoulder lifts systematically occur relative to the verbal responsive turn. We thereby illustrate how the actions to which shoulder lifts contribute are bound to (and derive their meaning from) specific turn positions. This is crucial for the different kinds of stance-taking that shoulder lifts can accomplish. The uses we show include shoulder lifts in pre-beginnings of responses (Section 5.1), lifts accompanying verbal responses (Section 5.2), lifts used before—and alternative to—projected turn completion (Section 5.3), lifts after a possible turn completion (Section 5.4), and shoulder lifts as stand-alone responses (Section 5.5). We present examples of both one-sided and two-sided shoulder lifts to build our argument that these two types of lifts have different affordances when it comes to stance-taking. By showing that shoulder lifts are a positionally sensitive resource for speakers in projecting, building, and retrospectively showing stances, we also hope to showcase the unique potential of conversation analytic and interactional linguistic approaches to further our understanding of multimodal stance-taking in interaction.

5.1 Shoulder lifts in pre-beginnings

We begin our analysis with shoulder lifts that occur in pre-beginnings, i.e., before a responsive turn's recognizable beginning. In this position, speakers use shoulder lifts to *project* the stance a responding speaker will take *before* they start of a verbal turn—a function that other non-verbal resources, such as frowns, smiles (Schegloff, 1996a; Kaukomaa et al., 2014), or sighing (Hoey, 2014; Schegloff, 1996a, pp. 105–106; see Section 2) can also accomplish.

Extract (2) is from an interaction between Isabell (ISA) and Ferdinand (FER) during a car ride. Before the extract, the participants had started talking about different rental care companies. Isabell now notes that two companies are represented throughout Europe and that she believes that one of them, SIXT, was founded in Germany. In line 2, she initiates a telling to support this belief: she saw a report about the wealthiest families on a TV channel (*RTL*) and the owners of SIXT were among them (lines 2–7). After a minimal acknowledgment by Ferdinand in line 9, Isabell reiterates that they belong to the richest families in Germany (line 11–12). She then downgrades her certainty regarding their country of origin (line 12), and in line 14, she settles on Germany as likely correct. After a 1.7-s pause, in which no reaction from Ferdinand is forthcoming, she starts moving her gaze toward Ferdinand (line 15).

Extract (2): FOLK_E_00301_SE_01_T_01_DF_01_sixt

https://bit.ly/3BJlg8P

```
01          (0.7)

02   ISA:   ich hatte ma nämlich so au(h)f: e(h)r tee el war das glaub ich
            (because) I once saw like on             RTL I think it was

03          so_n bericht über die reichsten familien gesehn.=weil die
            a report (you know) about the wealthiest families because they

04          gehen immer auf (irgendwie) so ne veranstaltung, .h
            always go to (like) a certain kind of event

05          (0.4)

06   ISA:   und da war nämlich die familie: (0.2) der sixt gehört auch
            and there was (in fact) the family that owns SIXT ((was)) also

07          dabei.=aber die sind >voll öffentlichkeitssche:u;<
            there but they are very private ((lit.: shy of the public))

08          (0.5)

09   FER:   kay.
            okay

10          (1.4)

11   ISA:   die gehören nämlich zu den reichsten familien
            because they belong to the richest families

12          deutschlands.=↑oder war das schweiz↑;
            in Germany     or was it Switzerland

13          (0.7)

14   ISA:   ich glaub es war deutsch#land#,
            I think it was Germany

15=>        (1.7)*(0.3)%(0.7)%(0.6)
     isa          *gaze tw FER-->
     fer                %raised eyebrows w/mouth shrug,gaze at ISA%
                              %L shoulder lift w/gaze away-->

16=> FER:   .h %(könnte) beides sein.
                (could) be either
     fer    -->%

17          (0.7)
```

Although Isabell displays doubt regarding her knowledge of the country of origin of the company's founders (lines 12, 14) and thus demonstrate s trouble with remembering, Ferdinand does not produce any responses. This may be why he seems to interpret Isabell's gaze toward him (line 15) as mobilizing a (lacking) response (Stivers and Rossano, 2010). As soon as Isabell starts turning her head and gaze toward him, Ferdinand raises his eyebrows, turns down the corners of his mouth, gazes directly at Isabell, and produces a left-shoulder lift right before suggesting that either country (Germany or Switzerland) is plausible.

In neither confirming nor disconfirming Isabell's assumption, Ferdinand produces a non-answer (Stivers, 2022, Ch. 3), which displays not only a lack of knowledge but also a certain degree of uncooperativeness by not telling what he thinks is likely. With his verbal and non-verbal behavior, Ferdinand observably refuses to engage in further reasoning about the matter in question and thus with the activity initiated by Isabell's trouble in remembering. Such one-sided shoulder lifts project a response that does not "go along" with the terms, expectations, or agenda set up by the prior speaker's turn, and they thus mark the speaker's resistance to engage with the Other's projectable response expectations, topical and sequential development, and the epistemic positioning expected from them. In doing so, responding speakers mark their response as non-expandable and propose sequence closing.

Non-answers, or misaligned responses, are not the only type of uncooperative responses projected by shoulder lifts in our data. The following extract, taken from a conversation between three friends, shows how a one-sided shoulder lift can be used for projecting a *disaffiliative* response. Melissa has just said that she and her partner

were considering moving in together and that she can imagine taking that step (lines 1–2). Saskia responds by expressing doubt about living together with her own partner, Thomas (lines 4–8) and then accounts for this by anticipating a specific challenge (lines 10–14), namely that her *sparsamkeitsverhalten* ("way of saving money;" line 14) may differ from his and that this may lead to friction. After a pause, Melissa produces a one-sided shoulder lift (line 15) and then a verbal response (line 16–19).

Extract (3): FOLK_E_228_SE_01_T_01_c1310_das ist doch normal

https://bit.ly/3BPigaQ

```
01    MEL:  (nee/nja) ich kann mir schon vorstellen
            I can indeed imagine

02          mit ihm zusammenzuziehn,
            moving in with him

03          (0.5)

04    SAS:  hm.

05          (0.3)

06    SAS:  !t .hh ich glaub net=also: äh ich hab mir das jetz noch
                   I don't think   so  uh now I haven't

07          nich so richtig überlegt ob ich mit_m thomas zusammenziehn.
            really thought about whether I want to move in with Thomas

08          (.) also ich mein d_s auch viel zu früh. .hhh
                so I mean it's also way too early

09          (0.1)

10    SAS:  aber ich glaub was (.) ein punkt (.) sein wird,
            but I think what       one point will be

11          (0.5)

12    SAS:  de:r mich beziehungsweise ihn stören wird, .h dass mein
            that for me or rather him will be annoying  ((is)) that my

13          putzver- (.) äh nich mein putzverhalten.=sondern
            cleani-      uhm not my cleaning behavior but rather

14          mein sparsamkeits+verhalten .h+ anders is+ als seins
            my way of saving money is different from his
      mel                   +head slightly up+down--+

15=>        (0.5)*(0.5)
      mel        *R shoulder lift->

16=>  MEL:  °ja*°=aber ich mein das is doch normal wenn man
             (yeah/well) but I mean that's PRT normal when one
      mel    ->*

17=>        *zusammenzieht-=dann (hat/muss) (.) *der an*dere
             moves in together then the other ((person)) has
      mel   *R shoulder lift--------------------*,,,,,,*

18=>        halt (macht/mal tun) (die dinge) (.) die dir nich passen?=
            (PRT) (does/PRT to do) (the things)    that don't suit you

19          =[(un dann)/(werden)]
              (an then)  (will)

20    SAS:   [ja.               ]
              yes

21          (0.5)

22    SAS:  ja deswegen muss du die jetz erscht ma kennenlernen. .h
            (yeah/well) that's why you've got to first identify them now
```

The pause that emerges after Saskia's turn (line 15) can already project a dispreferred response to the problem, or hypothetical trouble, that Saskia describes. In line 16, Melanie responds verbally and assesses the situation Saskia just described as *normal*. With her assessment, which indexes problems with the prior action's terms with a turn-initial *ja* ("yes/well," line 16; see Betz, 2017), Melanie casts Saskia's anticipated problems as a typical and expectable part of moving in together (rather than as out-of-the-ordinary, legitimate reason for *not* moving in together). In doing so, she pushes back against Saskia's displayed concerns. With the modal particle *doch*, she additionally appeals to common ground knowledge (Pittner, 2006) and thus to something that should be known to Saskia, which further contributes to rejecting the displayed problematicity of Saskia's concerns. All these features additionally contribute to Melanie's not just expressing her own subjective opinion on the matter but rather presenting her turn as an obvious fact known to, at least, the co-present interactants. In responding this way, Melanie curtails further development of this sequence into a troubles telling.

As in Extract (2), the shoulder lift in (3) precedes (and overlaps with the beginning of) a response that does not "go along" with the terms of the prior speaker's actions. However, while in Extract (2), a shoulder lift accompanied a *misaligning* response, it is used in (3) to project an upcoming (at least, partially) *disaffilaitive* response that rejects the relevance of Saskia's concerns. In prefacing her response with a shoulder lift, Melanie communicates resistance to engage in treating these concerns as expandable. The verbal turn that follows (lines 16–18) unpacks this positioning by framing Saskia's concerns as not out-of-the-ordinary but rather expectable (and therefore overblown here) and not expansion-worthy.

We also find *two-sided* shoulder lifts that occur in pre-beginnings of responsive turns. Extract (4), taken from the same fondue interaction as Extract (1), demonstrates this. Gero is telling Zoe and Norbert about his brother's first girlfriend, who their mother disliked. Norbert treats this information as difficult to reconcile with his own expectations of how Gero's mother would act. In response, Gero initially struggles to justify or explain his mother's behavior (lines 1–2). He then concedes that the situation was "weird" (line 6), which is supported by his description of the mother's reaction to the girlfriend's ultimate departure (line 9, see Pomerantz, 1986, on *extreme case formulations*). Zoe reacts with a short nasal laugh (line 11), while Norbert acknowledges the information with a minimal *okay* (Betz and Deppermann, 2021). After a pause, Zoe makes an assessment, *wie nett* ("how nice"). Accompanied by smiling and laughter, this assessment is intended as ironic and can be interpreted as casting either Gero's previous *formulation* of his mother's extreme reaction or the nature of *the reaction itself* (happiness) as insensitive or callous. In response, Gero raises his eyebrows, lifts both shoulders, opens his hands into a palm-up gesture, and states that he witnessed the situation first-hand (lines 14–15).

Extract (4): FOLK_E_00293_SE_01_T_03_c234_wie_nett
https://bit.ly/40VNNSy

01	GER:	hhh isch weiß net.=es war irgendwie .h isch denk
		I don't know it was somehow I think
02		dass es halt son
		that it was just (kind of/like) a
03		(0.2)
04	GER:	des war ja auch die erste wo (.) ins haus kam.
		(she) was also the first one who came to our house
05		(2.5)
06	GER:	un (.) ja,=es war irgendwie halt komisch, (.) aber mit der
		and yeah it was somehow just weird but with her
07		konnt meine mudder gar net.
		my mom couldn't get along at all
08		(1.9)
09	GER:	un war auch Ü:belst froh wie die weg war.
		and was also extremely happy when she was gone
10		(0.2)
11	ZOE:	mHN.
12	NOR:	oKAY,
13		(0.4)

```
14    ZOE:  wie nett.# *h h h h .hh#
            how nice
      ger             *palm-up gesture and two-sided shoulder lift-->
      fig           #10          #11

15    GER:  .pt(n)* °°ja°°# isch hab_s halt mitgekriegt hautnah. .h
                  yeah/well I witnessed it first-hand
      ger      -->*
      fig              #12
```

Figure 10

Figure 11

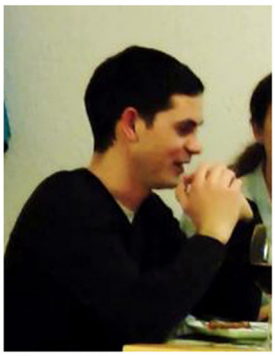


Figure 12

With his response in line 15, Gero provides evidence of his entitlement to produce the description in line 9. Unlike Extracts (2) and (3), a two-sided shoulder lift does not indicate the speaker's resistance to engage with the expectations set by the Other's action or activity. Instead, Gero's response accounts for his earlier turn (line 9), indicating that his knowledge is rooted in personal experience. In this context, the two-sided shoulder lift does not convey reluctance, unwillingness, or resistance to engage with Zoe's remark. Instead, it signals his inability to align with its implications [e.g., by recalibrating or reconsidering the stance he took previously, or by (co-)complaining about his mother] based on his experience. Gero's response to Zoe's negative, disapproving assessment frames his own prior statement as an objective recounting of events, rather than as a subjective or evaluative judgment. At the same time, with his turn in line 15, he positions himself as an observer with limited access to his mother's feelings, rather than as an active participant or evaluator of the situation. In doing so, he disclaims responsibility both for his mother's feelings and for his own description.

In this section, we have shown how shoulder lifts are used to project the type of the upcoming response and the stance taken in it. We have seen that one-sided shoulder lifts project responses that are in one way or another *uncooperative*, i.e., responses that do not "go along" with terms, expectations or agenda of the Other's prior action. In Extract (2), the recipient produces a misaligned and dispreferred response, namely a non-answer, which demonstrates a lack of willingness to engage with the prior search for an answer. In Extract (3), the recipient produces a disaffiliative response that rejects the problematicity of the Other's concerns and thus declines to affiliate with the prior speaker's stance. In both cases, speakers refuse to engage with, or commit to, the Other's initiated course of action, and the one-sided shoulder lift is crucial in conveying this stance. In contrast, the two-sided shoulder lifts in our data project responses in which speakers continue to orient themselves to the "social contract" between participants—namely, the moral responsibility of providing a cooperative response and re-establishing intersubjectivity (Extract 4). This is why two-sided shoulder lifts are of a more cooperative nature, as further sections will demonstrate. We have further shown how shoulder lifts are used to disclaim accountability, or responsibility, for what speakers are about to say by *distancing* from it, either by displaying a low epistemic (and deontic) stance and thus disclaiming the responsibility for the correctness of what is said (Extract 2 and 4) or by formulating the response as something self-evident (Extract 3).

5.2 Talk-accompanying uses of shoulder lifts

In addition to projecting a stance in a pre-beginning position, shoulder lifts can also index stance while accompanying verbal responses. Consider Extract (5) from a theoretical driving lesson. Before the extract begins, the instructor tells the student that she can opt to do license class C instead of C1. In lines 1–3, the instructor provides several advantages of this driving license category, which the student treats as news with *echt* ("really," line 4; Gubina and Betz, 2021). In response, the instructor confirms (line 5) and restarts her turn from line 3, in which she frames the mentioned advantages as a reason for her not understanding people who choose to do C1 instead (lines 8–9). The student then initiates a verbal response and, in parallel, raises her eyebrows, turns down the corners of her mouth (Figure 14), shakes her head and produces a two-sided shoulder lift (line 10; Figures 14, 15).

Extract (5): FOLK_E_00348_SE_01_T_02_DF_01_c320_cee_eins

https://bit.ly/3A6oLpf

```
01    INS:  weil  von      der grundgebühr,=un die fahrstunden,=
            cause in terms of the basic fee and the driving hours

02          =des is eigentlisch mehr oder weniger alles desselbe.=
             it's really more or less all the same

03          =des[halb ver]steh ich
             that's why I understand
      ins   >>gaze at STU--->

04    STU:      [echt,   ]
                really

05    INS:  ha ↑j[a,   ]
            of course

06    STU:       [ah j]a,
                  oh okay

07          (0.2)

08    INS:  deshalb versteh ich_s ja auch als net warum die leute
            that's why I also (generally) don't understand why people

09          cee #ei[ns machen.        ]=ne,
            (choose to) do C1, (y'know/right)?
      fig       #13

10=>  STU:          [*+.h *+%#ich§@(.)§]
                          I
      stu-f          *raised eyebrows*
      stu-g           +....+points at herself*palm-up gesture-->
      stu-b                 %two-sided shoulder lift-->
      stu-m                      §turns down the corners of her mouth§
      stu-h                       @headshake-->
      fig                   #14

11=>  STU:  ich hab mich da+%@ ehrlich gesagt# au nich das(s)
            I did              to be honest    also not that
      stu-g             -->+
      stu-b              -->%
      stu-h               -->@
      fig                                  #15
```

Figure 13

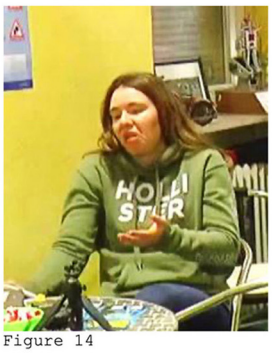


Figure 14

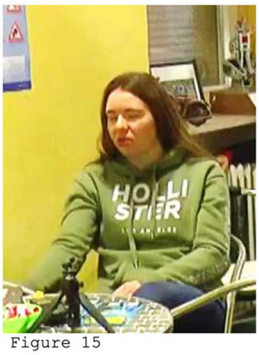


Figure 15

```
12=>        ich wusste gar nich dass es cee eins überhaupt gibt,
            I didn't know at all that C1 even existed

13          (0.6)

14    STU:  also
            so
```

```
15          (1.0)

16    STU:  die ham gesagt ja,=du machst (0.2) ce eins?
            they said   (well) you are doing C1

17          (0.4)

18    STU:  wenn ich gefragt hab was fürn führerschein ich brauch.=
            when I asked what kind of driver's license I need

19          =un dann hab ich gesagt #ja#.=#okay#. ((#=creaky voice))
             and then I said yes okay

20          (0.4)

21    STU:  #fertig#.
             done

22    INS:  .h h

23          (0.2)
```

This is one of the prototypical environments in which two-sided shoulder lifts occur in our data. The two-sided lift, which in this case is at turn beginning, is produced in response to a criticism-implicative action by the prior speaker: Since the student is part of the group of "people who do C1" instead of the C driving license, the instructor's lack of understanding (lines 8–9) concerns the addressed student as well and can be seen as an account solicitation. The student's account is not only produced with multiple embodied displays but also with a hedge (*ehrlich gesagt* "to be honest"), orienting to the prior as a challenge through "speaking sincerely" (see Clayman, 2010). The two-sided shoulder lift prefaces an account that conveys the speaker's *inability* to fulfill the Other's expectations, which can be inferred from the negative epistemic marker that displays the student's lack of knowledge (line 12). Thus, the student is claiming an inability to make any other choice, which amounts to positioning herself as not responsible and accountable for her choices. This also becomes observable in the following sequential context, when the student is reporting that she was simply told what kind of driving license she needed. In doing so, she disclaims agency and ascribes responsibility for the decision to others (apparently, to her employer).

A similar use of a two-sided shoulder lift is found in the case we showed in Section 1. We are returning to the point at which Zoe produces a criticism-implicative turn regarding Gero's basis for judging his female colleagues. Zoe's turn does not explicitly assess Gero's behavior, although its design (a *dass* "that"-clause) projects a main clause containing Zoe's evaluation. Nevertheless, stating that Gero is assessing colleagues (of the other gender) only by their appearance (rather than their abilities, professional competence, character etc.) is criticism-implicative (and face-threatening) because he is thus positioned as superficial and/or unprofessional. Furthermore, Zoe uses the extreme case formulations *immer* ("always") and *alle* ("all/everyone") to frame his behavior as a pattern (see Pomerantz, 1986), which can serve as grounds for reprimanding. Zoe's turn in lines 10–11 can therefore be interpreted as an accusation. In response, Gero produces an account (line 13): He is not familiar with the intern's professional competence and thus lacks alternative grounds for judging her.

Extract (6): FOLK_E_00293_SE_01_T_04_DF_01_c913_zweitrangig

http://bit.ly/3BS40hu

```
10    ZOE:  dass du die immer (alle)
            that you always

11          nach dei nach der ↑optik↑ ↓bewerten musst.
            have to assess them (all) by you- by their looks

12          (0.5)

13 => GER:  *ja ihr fAchlisch kompetenz kenn isch *net*.
            (well) her/their professional competence I'm not familiar with
      ger   *two-sided shoulder lift-------------*,,,*

14    NOR:  °nh nh nh°

15          (2.0)*(1.5)*(2.9)
      ger         *lifts both shoulders*

16 => GER:  *dies is* mir aber auch relativ egal,
             but it doesn't really matter to me anyway
      ger   *L shoulder lift*

17          (0.3)

18    ZOE:  hm hm, (.) wollt_s grad sagen. (.) is eher zweitrangig.=ne,
            hm_hm   was just gonna say that  it's rather secondary, right?

19          (1.3)
```

As in Extract (5), the two-sided shoulder lift accompanies a response to a criticism-implicative turn. Specifically, it is produced while the speaker accounts for his behavior by claiming inability to judge on other grounds due to Gero's lack of knowledge concerning the intern's professional competence (note also the turn-initial *ja* "yes/well," which can index problems with the prior action's terms in responding; Betz, 2017). This claimed lack of epistemic grounds for making such judgments allows for Gero's positioning as not responsible or accountable for the criticized behavior and proposes sequence closure. In contrast, the second account he provides in line 16 does not claim an inability to meet the Other's expectations; it instead rejects the *relevance* of the prior point and thus of the basis for the criticism. We argue that this is a typical sequential environment for a *one-sided* shoulder lift, which in this case serves to display the speaker's lack of interest and, more broadly, to index resistance (rather than inability) to engage with the course of action initiated by the prior speaker.

This section has shown that responding speakers use both types of shoulder lifts to disclaim accountability and responsibility ascribed to them by prior speakers' actions. We have also shown that one-sided and two-sided shoulder lifts can convey different stances and thus lead to different outcomes in similar interactional environments. Two-sided shoulder lifts seem to mark the speakers' lack of *ability* to "go along," or to further engage, with the Other's initiated or projectable course of action—due to a lack of knowledge (Extract 5, line 12; Extract 6, line 13). In contrast, one-sided shoulder lifts index some sort of *resistance* to "go along" with the Other's initiated or projectable course of action (Extract 6, line 16).

5.3 Shoulder lifts before possible turn completion

Shoulder lifts can also be produced at points at which (lexical-syntactic) completion is projectable. In such cases, they occur before a projected adjective or other assessing turn component in turns that embody an evaluative stance. The projected verbal component (and thus grammatical completion) is suspended and the (same) speaker produces a shoulder lift. Such shoulder lifts do not constitute completions of the incomplete turns but rather mark a stance toward the projected assessment.

We present three extracts to illustrate this use. The first comes from a job application training. The trainee (TRE) is describing to the trainer (TRR) and his assistant (AST) the quality of two recommendation letters he received from a past employer. *Zwei*(*er*) "two" (line 4) and *drei* "three" (line 7) denote overall grades, with "two" being a higher/better grade than "three" (and "one" being the best). After reporting the grade received for one letter, the trainee shifts to the description of *des andere* "the other one" (line 10). He had already reported earlier (not shown) that this report "read okay," conceding that it was not of the highest quality by comparing it to the most desireable assessment, an "A+." At line 12, he again initiates an assessment by uttering *(des andere) las sich* "(the other one) read." The syntactic structure of the turn projects a somewhat positive assessment (e.g., *gut* "good," *ganz gut* "pretty good"). However, instead of completing it, the trainee suspends the verbal turn in progress and produces a shoulder lift, a palm-up hand gesture (the latter in his lap and thus likely not visible to the trainer, see Figure 17), a slight mouth shrug, and a head shake (Figures 16–18).

Extract (7): FOLK_E_00173_SE_01_T_02_c669_las_sich_gut[4]

https://bit.ly/4gZ6paq

```
01    TRR:   °vielleicht ha[ben die en schlech]tes gewissen
             maybe they have a bad conscience

02    TRE:                [>also<-          ]
                           well/so
03    TRR:   und stellen ihnen trotzdem eins aus.° (äh/oder)
             and still issue one for you          (uh/or)

04           ↑zumindestens en sauberes zweier zeugnis↑.
             at least a clean grade two report

05           (0.6)

06    TRE:   nee ich glaub es wa:r äh zu unsere::r (.) vollen
             no I think it was     uh to our          full

07           zufriedenheit.=also [die <dr]ei.>
             satisfaction     so the three.

08    AST:                       [ hm_hm ]

09           (0.5)

10=> TRE:    des a[ndere wa]s ich gelesen *habe=
             the other thing I read was
      tre-f                               *eyebrow raise-->

11    TRR:         [ja      ]
                   yes
```

```
12=> TRE:   =+las %sich# &$.hh          &(0.3)* %#l+as sich %gu$&t.#$& %
              it read                      (it) read well
     tre-f                                -->*
     tre-h1       %slight head shakes-----------%          %head shake%
     tre-h2  +head tilt to L----------------------+
     tre-b              &two-sided lift&hold------------------&down&
     tre-g               $hands open,palm up-----------------$,,,,$
     fig             #16                      #17              #18

13          (0.3)

14   TRR:   °okay° .h (.) wenn sie_s ham schicken sie_s uns mal.=
             okay if you have it,       send it to us sometime
```

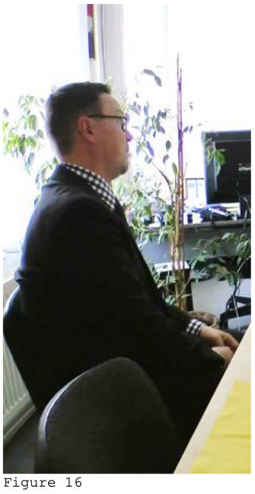


Figure 16

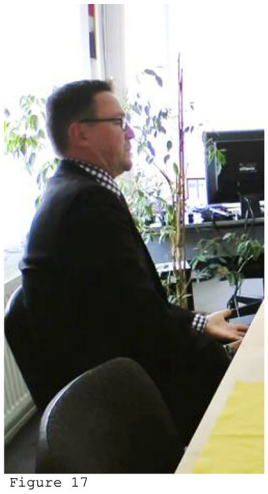


Figure 17

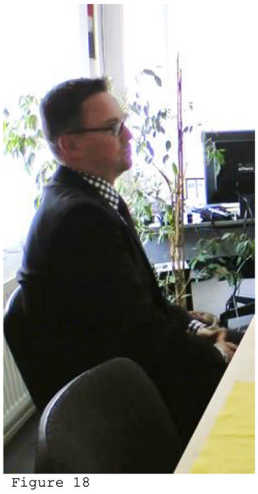


Figure 18

The two-sided shoulder lift is begun while the turn-in-progress (*las sich* "((it)) read"; line 12) is still incomplete. Specifically, the trainee's self-repair allows for the insertion of an embodied stance display. While the trainee is holding the gesture, he resumes his verbal turn and produces the projected assessment, *gut* ("good"). At its completion, he releases the shoulder lift. Interestingly, the shoulder does not constitute an embodied completion of the yet unfinished turn. Instead, it allows the speaker to display stance not with respect to a particular referent (the reference letter and its content) but rather with respect to his emerging positioning (Du Bois, 2007). The lift thus orients to the accountability for his assessment and specifically contributes to disarming, or disclaiming accountability: Since he did not get the best grade, a reference point he makes relevant in the prior context, this might imply that the reference letter is potentially critical or not that good, which contrasts with what he is saying in line 12. Furthermore, the trainee uses an impersonal format in formulating his assessment, which frames the assessment not as his personal opinion but rather as something that he is not accountable for. Thus, the turn design of his assessment allows him to distance himself from the claim of authorship and responsibility for the content of the reference letter being good.

We can also note that by marking the quality of the report beyond his control or influence, the trainee frames it as just being the way it is, which is why no further (sequential) expansion or elaboration is necessary. This is also what the headshake might be accomplishing in this turn, namely marking that "there is no need to discuss it" any further (Kendon, 2002, pp. 170–71). The trainer's reaction is responsive to this: His subsequent conduct suggests that he hears the trainee as communicating that he has nothing to add beyond what the trainee has arrived at with *gut*. With *okay*, the trainer accepts the trainee's description and moves to next steps (Mondada and Sorjonen, 2021): making arrangements for obtaining a copy of the letter in question (line 14).

However, not only the quality of the report, but also the *actual nature of the content* of the report is treated as beyond the trainee's control.[5] Given that in this professional context, reference letters are expected to avoid explicitly negative phrasing, the trainee may be understood to express awareness that the way in which his reference letter "reads" does not necessarily reflect the actual evaluation expressed by it. In view of the prior context not included here, the shoulder lift can be seen as a resource for disclaiming expertise in reference letters, and ascribing (relatively more) expertise to the trainer, thus leaving him with the responsibility to ascertain the actual upshot of the reference letter. The request in line 14 may reflect that, having been designated the expert role, the trainer now knows he is responsible for reading the report himself to get an accurate picture of its contents, and he thus asks for a copy.

Recipients of such (syntactically) incomplete turns with shoulder lifts before a possible turn completion orient to those shoulder lifts as projecting a particular stance in an embodied way. This is visible in instances in which others treat the verbally incomplete turn as *complete*, for example, by responding with an agreeing second assessment. Extract (8), which comes from a theoretical driving lesson, illustrates this. The instructor is

4 Abbreviations used in the focus lines: Tre-f: face; Tre-h1: head1; Tre-h2: head 2; Tre-b: body; Tre-g: gesture.

5 We thank an anonymous reviewer for prompting us to make this dimension more explicit.

explaining the "BF 17" program for accompanied driving and the requirements for drivers who accompany students under 18 (which is the legal driving age in Germany). In lines 1–2, she criticizes the guidelines for not being specific enough, and a student agrees (line 4). After elaborating on this jointly with the student (lines 6–17), the instructor begins formulating a summary assessment, which includes elements that downgrade its strength: *also des is halt_n bissen* "so that's just a bit" (line 19).

Extract (8): FOLK_E_00348_SE_01_T_02_c153_wischi_waschi

https://bit.ly/4hfPwIT

```
01    INS:  allerdings steht nirgendswo geschrieben wo zum beispiel die
            however ((it)) is stated nowhere      where for example the

02          begleitperson sitzen soll.
            accompanying person should sit

03          (0.3)

04    STU:  ja, des find ich auch blöd.
            yes I think that's stupid too

05          (0.4)

06    INS:  also wenn der jetz hinter dir sitzt.=[un zeitung ] liest oder
            so if he is now sitting behind you and reading the newspaper or

07    STU:                                      [kofferraum.]
                                                 trunk

08    INS:  äh ja kö kofferraum geht eigentlisch net.=weil [vom        ]
            uh yes tr trunk is not really possible because of

09    STU:                                                 [ja doch.=we]nn
                                                          well yes it is if

10          man so_n siebensitzer hat.
            if you have a seven-seater

11          (0.8)

12    STU:  geht koffer[raum schon.]
            trunk is possible

13    INS:             [ja       ok]ay.
                        yes       okay

14          (0.4)

15    INS:  ja; (.) dann (.) ja. (.) hintn_
            yes     then     yes     in the back

16          (0.2)

17    INS:  klar.
            of course

18          (0.3)

19=>  INS:  also das is halt_n bissen#
            so  that just is a bit
      fig                          #19

20=>        *#(0.3)*#
      ins   *two-sided shoulder lift*
      fig    #20    #21

21    STU:  [$des find] ich auch %b[löd.$  ]
              I think that's stupid too
      stu-f  $frowns--------------------$
      stu-h                       %shakes head->

22    INS:  [.h       ]

23    INS:                          [des is ] alles so wischiwaschi%
                                       it's     all so wishy-washy
      stu-h                                                    ->%

24          was die mittlerweile machen.
            what they are doing now
```

Figure 19

Figure 20

Figure 21

Similar to Extract (7), the speaker does not bring her turn to its projected syntactic completion, which would, in this case, require an item that completes a negative assessment (of the guidelines she is supposed to teach) and can serve as a gloss for the situation instructor and student just expanded on (e.g., an adjective such as *unklar* "unclear," *vage* "vague" or *blöd* "stupid," see line 4). Yet, in contrast to Extract (7), in which the trainee himself completes the turn, even before lowering his shoulders again, the shoulder lift in this case occurs in a designedly incomplete turn, or an aposiopesis (Imo, 2011). When the adjective (or other negative evaluative term) is due, the instructor suspends talk and produces a two-sided shoulder lift (Figures 19–21). The sequence now develops differently than in the previous extract: In this case, the student responds before the turn completion by the instructor, and her response displays her understanding of where the prior speaker was going: She produces a negative assessment and marks it as an agreement with the instructor's stance (see *auch* "too," line 21). In addition to formulating an agreeing second assessment, she produces a head shake, which marks her (affiliative) disapproving stance (Kendon, 2002). As in Extract (7), the shoulder lift does not *complete* the syntactically incomplete verbal turn. It does not communicate an evaluation itself but rather a stance toward the *projected* assessment, that of the lack of necessity to go on record with the word. With the lift, she orients to accountability for her position: The projected assessment is framed as obvious (so obvious it does not need to be put in words; and avoiding putting it into words may also be motivated by her role as a driving instructor teaching the rules that are being assessed negatively here) and as thus not worthy of elaboration.

Shoulder movements placed at points, at which (summary) assessments or upshots (e.g., with *also* "so" or *aber* "but") are projected but not produced, can also be realized as one-sided lifts. With these, interactants emphasize the inescapability of the projected summary/conclusion as well as its obviousness. In Extract (9), the recognizability of what is left unsaid, and the grounds for treating it as obvious emerge from the larger sequential context. Elena and Norbert are newlyweds, and we join them in a conversation that takes place one week after their wedding. Earlier Elena had assessed time they spent with a friend on the preceding day as *schön* ("nice"), and she also reported talking to this friend on the phone and helping her identify guests on photos from the wedding. About 40 min later the exchange shown below happens. After a minimal embodied response by Norbert (line 13) to Elena's assessment *det sind so richtig coole freunde* ("they are just really great friends;" line 13), Elena continues with a specification, or a specifying account, a summary assessment (line 15), and a shoulder lift (line 17).

Extract (9): FOLK_E_00039_SE_01_T_02_c1255_coole_freunde

https://bit.ly/4eG4aHo

01	ELE:	ah ick fInd det schö:n. *oh I find it nice*
02	NOR:	ja wa, *yes right*
03		(0.6)
04	ELE:	.pt jo:. *yes*
05		(2.2)
06	NOR:	.hhh
07	ELE:	nee mit den beeden da fühl ick mich ooch richtig wohl_ *no with the two of them I really feel comfortable too*
08		(0.7)
09	ELE:	flori und karin

```
10    NOR:    (.) ↑↓jo:,
                  yes

11    ELE:    (.) >JO.<
                   yes

12            (0.9)

13    ELE:    det sind so richtich (0.6) coole freunde. *(.) chHH        *
              they are just really cool friends
      nor                                              *nods+gaze at ELE*

14            *(6.6)
      nor     *gaze away from ELE-->

15=> ELE:     da +brauch+_ma o  +jar +nischt groß%artig +sagen+.=dit [is  ]so:
              you don't need to say anything big about that that is so
      ele-b      +two-sided lift+lift+                  +lift+
      ele-g                                      %gaze at NOR-->

16    NOR:                                                          [nee.]
                                                                     no

17            +%(0.3)%+
      ele-b   +R shoulder lift+
      ele-g    %,,,,,%gaze away from NOR-->>

18            (1.9)*(0.8)
      nor       -->*gaze to ELE-->>

19    NOR:    (na/ne) wissen bescheid.
              (well/no) they know what's up

20    ELE:    (.) die wissen bescheid +>ja_<+
                  they know what's up yes
      ele-b                          +R shoulder lift+
21            *(0.2)
      ele-b   *nods,w/decreasing amplitude-->>
```

In line 15, Elena describes a positive aspect of the friendship she just assessed positively (unspoken understanding), thus also accounting for her assessment. Furthermore, Elena's turn in line 15 is (at least partially) responsive to Norbert's lack of a response, given that her own prior turn (line 13) makes a second assessment or (dis)agreement expectable. In a second TCU in line 15 (and in overlap with Norbert's agreement with *nee* "no," line 16) she begins to produce a projectable assessment of how she feels about the friendship with *dit is so:* "it's like/that is so" but does not produce the projected descriptor.[6] Instead, she lifts her right shoulder. Similar to Extracts (7) and (8), rather than constituting an embodied completion of the projected assessment term (cf. Skogmyr Marian, 2021), a shoulder lift in this position conveys a particular stance toward the projected assessment, namely that it is obvious, or self-evident. This is also what Elena conveys with her first TCU in line 15, and it works toward disclaiming the speaker's own accountability and responsibility for the assessment (due to its self-evident nature). This obviousness of the assessment contributes to the meaning of the lacking necessity to expand any further, which is additionally highlighted by Elena's gaze shift away from Norbert (line 17). Norbert, however, now shifts his gaze *to* Elena (line 18) and offers a formulation that demonstrates his understanding of where Elena's turn was headed as well as agreement with her assessment (see Sacks, 1992, vol. I, pp. 146–47 on demonstrating vs. claiming understanding). Elena responds to this with a confirming repeat turn, indexing that Norbert has restated what she had already alluded to (see Schegloff, 1996b) and in her confirmation she produces another shoulder lift on the confirmation particle (see Section 4.1).

Extracts (7)–(9) have demonstrated how shoulder lifts contribute to stance management at points at which completion is projectable, more specifically before a projected adjective or other assessing turn component that display an evaluative stance. In such cases, the shoulder lift is produced when the projected grammatical completion is suspended. We have shown that such shoulder lifts *do not complete* the suspended turns but rather *mark a stance toward the projected* assessment. In particular, shoulder lifts can project an evaluative stance that can be problematic due to contrasting expectations and expertise (Extract 7) or due to the obviousness and self-evident nature of the upcoming and projected stance display (Extracts 8, 9). Furthermore, we have observed that shoulder lifts can occur either within a momentary, repair-like suspension before the speaker completes their turn (Extract 7), or with a designedly unfinished turn, after which no completion of the turn follows and another speaker takes over (Extracts 8, 9). This distinction arises because in Extracts 8 and 9, with

6 The incompleteness of Elena's TCU could also be accounted for by Norbert's early affiliative response *nee* ("no", line 17).

aposiopesis, the turn end (i.e., what will complete the assessment) is clearly projectable for the interlocutor. In contrast, in Extract 7, the trainer cannot predict exactly what will follow.

5.4 Shoulder lifts in post-possible completion position

In Section 5.3, we presented shoulder lifts that occur before a possible turn completion. We have demonstrated how such lifts can project a stance toward an upcoming assessment or other evaluative descriptor before its production. We now turn to shoulder lifts that are produced in a post-possible completion position and show how these are regularly used as *post-completion* stance markers. "Post-possible completion" of a turn is a recurrent and systematic place for not just verbal turn extensions but also for elements that constitute "retrospective or retroactive alignments toward it, or consequences of it" (Schegloff, 1996a, p. 90), i.e., retrospectively taking a stance toward prior talk and what can be inferred from it. Such "post-completion stance markers" (ibid) can include grammaticalized elements, for example stance-marking particles in certain languages, and other linguistic and also non-vocal resources—"post-completion nodding, facial expressions (e.g., smiles or grimaces), shrugs, posture shifts, disclaimers ('I dunno'), laugh tokens, coughs, exhalations and sighs, in-breaths, and I know not what else" (Schegloff, 1996a, p. 92; see also p. 121, note 37). In our collection, post-possible completion shoulder lifts are typically produced as two-sided lifts. In contrast, one-sided shoulder lifts are rare and similar to uses of shoulder lifts that precede and accompany talk, i.e., they communicate a stance of resistance rather than a stance of inability to do/say/add more (see Sections 5.1 and 5.2). In this section, we present two examples of two-sided lifts.

Our first case comes from the theoretical driving lesson that we already showed in Extract (5). In Extract (5), we focused on the student's account for choosing the driving class C1 instead of C: a lack of knowledge (lines 10–12). Note that the teacher slightly nods at this point (line 13) but continues gazing at the student. This can be interpreted as treating the student's account as incomplete/insufficient at this point and inviting further elaboration.

Extract (10): FOLK_E_00348_SE_01_T_02_DF_01_c325_cee_eins
https://bit.ly/3A6oLpf

```
07          (0.2)
      ins-g >>gaze at STU-->>

08    INS:  deshalb versteh ich_s ja auch als net warum die leute
            that's why I also (generally) don't understand why people

09          cee ei[ns machen.]=ne,
            (choose to) do C1, (y'know/right)?

10    STU:        [.h ich (.)]

11    STU:  ich hab mich da ehrlich gesagt au nich das(s)
            I did          to be honest   also not that

12          ich wusste gar nich dass es cee eins überhaupt gibt,
            I didn't know at all that C1 even existed

13          ~(0.6)~
      ins-b ~nods-~

14    STU:  %also%#
             so
      stu-b %two-sided shoulder lift%
      fig         #22

15=>        %#(1.0)%
      stu-b %x2 two-sided shoulder lift%
      fig    #23

16    STU:  #die ham gesagt ja,=du machst (0.2) ce eins?
             they said (well)   you are doing   C1
      fig   #24
```

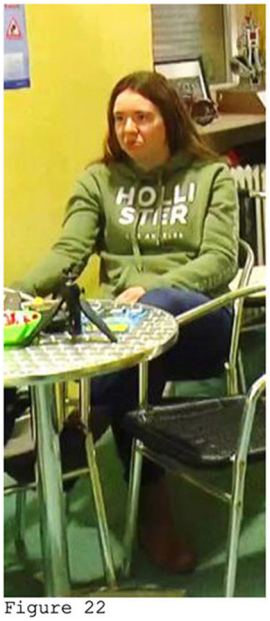


Figure 22

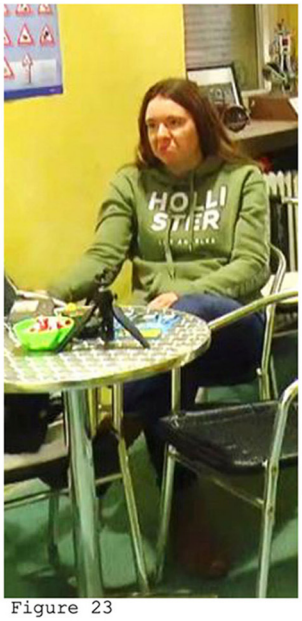


Figure 23

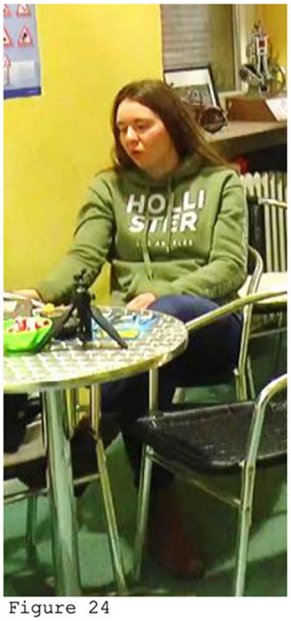


Figure 24

17 (0.4)

18 STU: wenn ich gefragt hab was für_n führerschein ich brauch.=
when I asked what kind of driver's license I need

19 =un dann hab ich gesagt #ja#.=#okay#. ((#=creaky voice))
and then I said yes okay

20 (0.4)

21 STU: #fertig#.
done

22 INS: .h h

23 (0.2)

After the pause in line 13, the student produces *also* ("so"), which, similarly to the stand-alone *so* in English, "can be deployed to project an unstated upshot after a prior turn has been brought to a possible completion, and some silence begins to emerge, or after a recipient has produced some (minimal) uptake of that prior turn" (Raymond, 2004, p. 192; cf. Alm, 2015). Following this, the student performs two two-sided shoulder lifts while looking at the teacher (Figure 23), still refraining from continuing her turn. We argue that with these post-completion lifts, the student indexes her inability to further account or elaborate, thus disclaiming accountability for her choice of the driving lesson. While the student's interactional conduct in lines 13–15 can be interpreted as an attempt to invite sequence closure, the teacher's conduct does not align with this course of action, as evidenced by their lack of verbal response and continued gaze.

A similar case comes from a board game interaction. Christine, Renate, and Gabriele have just sat down; the fourth player Vanessa, who organized the game, set up the recording, and is also the most experienced player, is not yet seated. In her absence, Christine is distributing game pieces, and Renate is explaining aspects of the game to the novice player Gabriele (lines 7–10). In 12–15, Christine suggests that Gabriele change seats, which the latter resists in lines 16/18. Christine then proposes reseating herself (line 21). This is met with a rejection by Gabriele (*naa* "no," line 23). The account that follows (reissued from line 16) suggests that the absent participant arranged the seating purposefully (*extra* "on purpose"), thus implying that the chair Christine proposes to fill was left empty by design. In formulating her account in this way, Gabriele ascribes the responsibility for the seating arrangement to Vanessa and thus disclaims responsibility and accountability for rejecting Christine's proposal. Instead of acknowledging Gabriele's account, Christine continues gazing at her. We are interested in Gabriele's bodily conduct at this moment (line 24).

Extract (11): FOLK_E_00357_SE_01_T_01_c64_extra_dahergesetzt

https://bit.ly/4h2vUaP

```
01    CHR:   da brauchen wir uns jetz net groß drum zu kümmern?
             we don't need to bother much with that right now

02    REN:   [des soll dan]n .h [die (.)    vanessa] °machen,°
              that's something Vanessa should take care of then

03           [genau.      ]     [die vanessa machen]
              exactly           Vanessa (should) take care of

04    REN:   des kriegste da[nn auch mit der ] zeit mit;
             you'll also catch on to that over time

05    CHR:                  [(hier)          ]
                              here
06           (0.3)

07    REN:   also wenn [(.) des hab i]ch dir glaub ich schon mal kurz=
             so if          I think I've already told you briefly

08    CHR:             [(eh/wie/die) ]
                       eh/how/what

09    REN:   =erzählt, wenn d- keiner deinen tit- deine karde
                      if n- nobody guesses your titl- your card

10           [errä:t dann kriegs][te kein: ][punkt.      ]
                    then you don't get any points

11    GAB:   [hmhm .hmhm°       ]

12    CHR:                       [.h       ][dann könnte ](mer/ma) des:
                                              then (one/we) could (do) it

13           (0.8)

14    CHR:   .h we_mer vielleischt (.) wenn du disch hier
                if we perhaps          if you sit down here

15           hinsetz[sch (sie kann dann)]
                     then she can

16    GAB:          [ja sie hot mich jet]z (.) extra da [v(h)ielleicht]=
                     (well) now she has explicitly here-         maybe

17    REN:                                                [jA.          ]
                                                           yes

18    GAB:   =setzt sie sich da hin.
              she will sit down there

19           (1.1)

20    CHR:   ja_a,
             yea

21           weil ich (0.2) oder isch setz misch do <owwe hin_>
             because I      or I will sit over here

22           .h [dass SIE           ]
                so that she

23=> GAB:       [*naa,+sie hot mich+ E]Xtra da her#g++set#zt.#
                 no she explicitly seated me here
     fig                                        #25    #26 #27
     chr         *gaze at GAB-->
     gab              +,,,,,,,,,,,,+upp body back,
     gab                           opens hands/arms++fwd-back mv w/R hand->
```

Figure 25

Figure 26

Figure 27

24=> REN: **$jaja,+%di#e va#nessa ha#[t]#%+**
yesyes Vanessa did-
fig #28 #29 #30 #31
gab ->+arms widen,**two-sided shoulder lift**+
gab %slight head shakes----%
ren $gaze at CHR-->

25 CHR: [˘j]A:,*$(.)dass sie des dann# macht.=
yea (so that) she should do it then
chr -->*to game->>
ren -->$

Figure 28

Figure 29

Figure 30

Figure 31

After rejecting Christine's proposal and accounting for it (line 23), Gabriele's upper body moves slightly backward (see Figures 25, 26), while opening her arms and hands into a palm-up gesture (see Figures 28–30; Kendon, 2004; Müller, 2004). While Christine continues gazing at Gabriele and produces no response (lines 23–24), Gabriele lifts both shoulders (see Figures 29–31) and produces slight lateral head shakes before retracting the gesture (see Figure 8). This embodied conduct, which emerges simultaneously with Renate's confirmation in line 24, retrospectively contextualizes her verbal account in line 23: She conveys an inability to offer more than this account, which concerns the intentions of an absent third person. We can note that Gabriele's shoulder lift is done *after* the possible end of her verbal turn and in the face of Christine's continued gaze on her. This gaze may convey a continued expectation of further accounting for her rejection (Stivers and Rossano, 2010). With a post-completion two-shoulder lift, Gabriele communicates a particular stance of inability to further explicate, or elaborate, thus disclaiming accountability for the arranged seating and for not having more to say (which, as discussed above, she already accomplishes with her verbal account in line 23) and proposing sequence closure (see also Hoey, 2014: p. 184–5 on similar functions of post-completion sighs). In contrast to Extract (10), Christine here accepts this proposal by receipting Gabriele's response (line 25) and also performing an embodied shift to game objects.

In this section, we have demonstrated how shoulder lifts are used in post-possible completion position. Such shoulder lifts, which in this position typically occur as two-sided shoulder movements, seem to push back against the Other's expectations of more to come (which are usually occasioned and displayed by continued gaze and/or focus of the Other toward the recipient) as well as reject further accountability. One-sided shoulder lifts,

by contrast, can function as stance markers that can retrospectively frame assessments (e.g., downplay their reach or consequentiality in the here and now) and thus can help speakers manage face-threatening aspects of their prior action.

5.5 Shoulder lifts as stand-alone responses

In the previous sections we have seen how shoulder lifts are connected to (precede, co-occur, follow) talk that elaborates or restricts the movement's import. But shoulder lifts can also be used as stand-alone responses to convey resistance to another speaker's proposed course of action. In other words, shoulder lifts can be complete and coherent actions in themselves, as evidenced by participant orientations to them. In these uses, it is the ongoing course of action and the relevancies and expectations established by the prior turn that allow the shoulder lift to take on a particular meaning and convey a specific responsive action.

We have already seen (although not yet discussed) such a shoulder lift in our initial extract, to which we now return. Recall that Zoe had produced a criticism-implicative turn, targeting Gero's criteria for judging women. In response to what can be heard as an accusation, Gero produces an account (line 13, see detailed analysis of this in Section 5.2). He claims a lack of knowledge (i.e., an inability to use a better/more acceptable basis for his judgement) as a defense. Zoe does not respond to this; a silence emerges, in which Zoe continues eating and then shifts her gaze to Gero. Following this, Gero lifts both shoulders (line 15).

Extract (12): FOLK_E_00293_SE_01_T_04_DF_01_c913_zweitrangig
http://bit.ly/3BS40hu

```
01    GER:  ((smacks)) wir ham jetz auch ne prak- ne (.)
                       we now also have an int- a

02          brasilianische praktikantin.
            Brazilian intern

03    NOR:  ja, (.) und?
            yes     and

04          (0.3)

05    GER:  pf:_

06          (0.83)

07    NOR:  nh nh nh nh nh ((laugh through nose))

08    GER:  voll hässlisch.
            really ugly

09    NOR:  ((laughs for 1.7 sec.))

10    ZOE:  dass du die immer (alle)
            that you always

11          nach dei nach der ↑optik↑ ↓bewerten musst.
            have to assess them (all) by you- by their looks

12          (0.5)

13    GER:  *ja ihr fAchlisch kompetenz kenn isch *net*.
            (well) her/their professional competence I'm not familiar with
      ger   *two-sided shoulder lift--------------*,,,*

14    NOR:  °nh nh nh°

15=>        $(0.5)%(1.5)$*(1.5)                    *$(0.2)%(2.7)
      ger-g $gaze at ZOE$turns head away-----------$
      zoe         %gaze at GER--------------------------%
      ger-b               *two-sided shoulder lift,
                           raised brows and palms up*

16    GER:  *die is* mir aber auch relativ egal,
             but it doesn't really matter to me anyway
      ger   *L shoulder lift*

17          (0.3)

18    ZOE:  hm hm, (.) wollt_s grad sagen. (.) is eher.=ne,
            hm_hm   was just gonna say that  it's rather secondary, right?

19          (1.3)
```

Gero's two-sided shoulder lift is produced after he has finished an account (line 13). It is, however, not designed as a post-completion stance marker. Note that Zoe shifts her gaze to Gero in line 15, which might be indexing the continued relevance of accounting. Gero seems to respond to this "requesting" gaze with a shoulder lift. This embodied response disclaims (further) accountability and retrospectively frames his verbal account as inescapable and non-expandable. He has already produced an account for the criticized behavior and (thus) has nothing else to add in his defense.

In Extract (13), the responding speaker conveys a refusal/declining of the opportunity to engage with a proposed action (assessing) via a shoulder lift, and it is oriented to as such. Ferdinand and Isabell are driving on the highway. Isabell has just checked the projected arrival time and assessed it negatively (i.e., as later than expected). Line 1 could be proposing a (non-serious) solution to this "problem": Driving faster will get them there earlier. In line 4, Ferdinand shifts his gaze to where one of the cameras is mounted, and in line 5, he formulates a noticing (while Isabell seems to be reading a highway sign announcing a landmark or upcoming regional attraction: *rock pop museum*, lines 6–7). Isabell then shifts her gaze to the same spot and confirms Ferdinand's observation (line 7). In overlap, Ferdinand offers a revised understanding of an action they did earlier (line 8), which can be seen as troubleshooting, and he assesses the whole situation as *nervig* "annoying" (line 9). Isabell responds with a shoulder lift (line 10, Figures 33, 34).

Extract (13): FOLK_E_301_SE_01_T_02_c116_is_ja_nervich
https://tinyurl.com/4x8jacx9

```
01   ISA:   gib ma_n bisschen gas gleich_
            step on the gas a bit now

02          (0.2)

03   FER:   nh ((inbreath through nose))

04          (4.2)*(0.6)
     fer         *gz fwd, slightly R, to
                  front cam(?)-->

05   FER:   .hh ja [%jetz ]läuft das% ne,
               well now it is working huh?
                   %R hand up/fwd--%

06   ISA:          [rock  ]
                    Rock

07   ISA:   (pop museum). [ja_         ]
            pop museum yes

08   FER:                [dann war das] viel*leicht doch falsch;
                         then maybe that was wrong after all
     fer                                  --->*

09   FER:   o+ah *is+         %ja      ner%*+*vig.#  +
            o:h (this) is PTC annoying
                             %L hand drop%
                  *head tilt twd ISA----------*turn,gz to ISA--->
     isa     +......+head turn/gz to FER----+away/fwd+
     fig                                          #32

10=>        (0.2)*&        #(0.2)*&#(0.6) *(0.4)* (2.6)
     isa          &L shoulder lift&
     fer     --->*gz away/fwd----*        *gazes L (>rear mirror?)*
     fig                  #33     #34
```

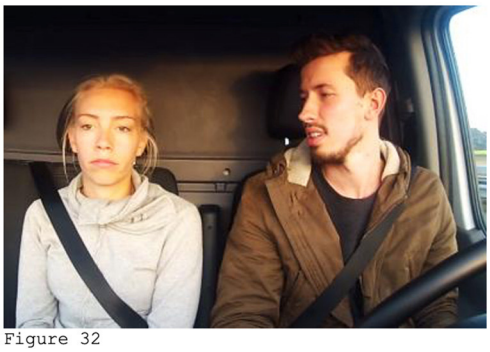


Figure 32

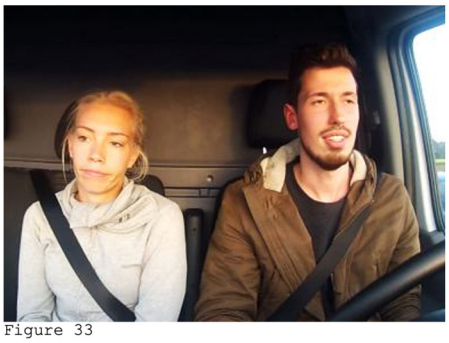


Figure 33

Figure 34

```
11   FER:   !t naja.
              well

12          +(0.2)+(0.5)
     isa    +.....+gaze twd R (window/cam)->

13   FER:   .h un %DIE is schon  wieder% abge+sackt.
              and this one has already dropped again.
     fer         %points w R hand to R%

14          (1.2) ((FER leans twd cam to her right,inspects it))

15   ISA:   !t filmt aber noch die straße;
              but it's still filming the road

16          (0.2)

17   FER:   ja,
            yes

18          (.)

19   ISA:   >(ja.)<
              yes

20          (1.5)
```

By expressing disappointment or regret in line 9 (notably with *oah*, Couper-Kuhlen, 2009; Golato, 2012), Ferdinand not only invites a second assessment but also offers an opportunity for Isabell to co-complain, or, at least, to affiliate with him. After Ferdinand directs his gaze to Isabell (mobilizing response; see Figure 32), Isabell produces a one-sided shoulder lift (with a head tilt and mouth shrug) as a response. Although Ferdinand is shifting his gaze away from Isabell as she produces the lift, her movement—done with her left shoulder, which is close to him—is likely still perceivable to Ferdinand. He responds with disengagement (gaze to left car mirror, line 10), a click, which can close down a current topic (Ogden, 2013), and a stand-alone *naja* ("oh well," line 11), which marks readiness to shift the topic while passing on the opportunity to expand or nominate a next one (Golato, 2018, pp. 420–22). Ferdinand then moves on to an observation about the other camera, and this is taken up by Isabell.

This extract provides sequential evidence for a co-participant's understanding/treatment of a stand-alone shoulder lift in responsive position. Isabell does not just decline to offer a verbal response but takes a particular stance toward Ferdinand's prior action: She declines to engage in the line of action Ferdinand offered in line 9 (co-complaining) by treating it as not worthy of expansion (or of any verbal response at all). Thus, like Extract (12), the shoulder lift here functions as a sequence exit device, which is oriented to by Ferdinand with a topic shift.

6 Conclusion and discussion

This study has focused on the use of one-sided and two-sided shoulder lifts in German talk-in-interaction and the interactional work they can accomplish in various turn positions. Focusing on broadly responsive uses, we have demonstrated how shoulder lifts are used in four major turn positions in our data, namely in pre-beginnings of the turn, accompanying the turn, at points at which completion is projectable and in a post-possible completion position. Shoulder lifts that occur in pre-beginnings typically project the uncooperative (misaligning or disaffiliative) nature of the upcoming action. Talk-accompanying shoulder lifts usually contribute to the uncooperative stance embodied by the action of the verbal turn. Shoulder lifts that occur before a possible turn-completion typically pre-empt projected evaluative terms and thus show speakers' orientation to the possible problematicity or non-necessity of producing the turn completion verbally. Finally, shoulder lifts in post-possible completion position mark the completeness and non-expandability of what was said before and thus push back against possible expectations of the Other for more to come. To summarize, our study elaborates the contribution of precise sequential placement within TCU and beyond to the function of a bodily resource or movement (Schegloff, 1996a). Thus, these results contribute to recent research analyzing the positionally-sensitive nature of bodily resources and sound objects in talk-in-

interaction (e.g., Clift, 2021; Ford et al., 2012; Hoey, 2014, 2020; Ogden, 2013).

We have also demonstrated that one-sided and two-sided shoulder lifts seem to contribute to accomplishing different interactional work. Specifically, two-sided shoulder lifts regularly mark speakers' lack of *ability* to "go along," or to further engage with, the Other's initiated or projectable course of action, often due to a lack of knowledge. Hence, such shoulder lifts could be seen as indexing low epistemic stance. In contrast, one-sided shoulder lifts tend to mark resistance to "go along," or to engage with, the Other's initiated or projectable course of action. In doing so, they aim to curtail the Other's line of action. Thus, although both kinds of shoulder lifts could be seen as uncooperative in some way, two-sided shoulder lifts still work toward re-establishing intersubjectivity and thus seem to be *relatively more cooperative* than one-sided shoulder movements.

Our research contributes to a growing body of work that focuses on the *division of labor* among variants of specific embodied resources (e.g., one- vs. two-sided), or how the shape/quality of a resource (e.g., small vs. conspicuous movement) can affect the functional spectrum of that resource (cf. Helmer et al., 2021 on downward vs. upward head nods; Hömke et al., 2017 on short vs. long blinks; Debras and Cienki, 2012 on head tilt left vs. right). This clearly demonstrates "order at all points" (Sacks, 1984, 1992) in the situatedness of various resources in interaction, specifically in the meaning that interactants derive from how talk, bodily resources, and sequential placement shape and particularize meaning together. As our analyses show, shoulder lifts also systematically co-occur with other body movements, such as head tilts, eyebrow raises, or palm-up orientations of the hands. Future research is needed in order to clarify to what extent these other movements matter for the interpretation of the shoulder lifts or are perhaps typical of particular uses of shoulder lifts.

We have also been able to determine the main job of shoulder lifts across different uses and contexts, i.e., a possible *context-free meaning*, namely indexing disclaiming accountability or responsibility for the projected or produced turn-at-talk. Our results show that there are different types of accountability that can be indexed as reduced with shoulder lifts: First, in cases in which they are produced in response to account solicitations or challenges, shoulder lifts work toward disclaiming *moral* accountability (see Extract 4, 5, or 6). Second, lifts can be used to push back against being made accountable for (not) producing a specific response and (not) following the normative expectations set by the Other's prior action (e.g., Extract 2, 3, 11, or 12). Finally, shoulder lifts can function in an anticipatory manner: Speakers can produce them to *pre-empt* something that might make them accountable (e.g., Extract 7, 8). Thus, this research presents new findings concerning the role of embodied resources in managing accountability and responsibility in talk-in-interaction (see Heller, 2018; Robinson, 2016).

Our findings also contribute to the growing scholarship on embodied resources in stance-taking in talk-in-interaction. In particular, we have shown that in contrast to many other bodily movements that can embody particular stances in social interaction (like iconic gestures or specific facial expressions, e.g., Kendon, 2004; Skogmyr Marian, 2021; Streeck, 2009), shoulder lifts do not take a stance toward the referent of their verbal turn nor toward the referent of the Other's prior turn. We have further shown that shoulder lifts are not connected with negative or positive stances, or connotations. Instead, shoulder lifts operate on another level of stance—stance toward what the speakers themselves are saying (or are about to say or just said) as self-evident, inescapable, obvious, and (thus) not further expandable or negotiable.

While one-sided and two-sided shoulder lifts in German seem to accomplish different interactional work, more cross-cultural and cross-linguistic research is needed to examine whether a similar distinction can be found in other cultures and languages as well. Furthermore, as we mentioned in Section 4, we excluded cases of shoulder lifts occurring in multi-unit turns from the current analysis. Thus, it would be important to examine such occurrences of shoulder lifts from a micro-analytic perspective in future studies. Finally, our work helps specify the "variety of meanings" (Debras, 2017, p. 29) for components of the "shrug," thus expanding our understanding of the context-specific use of the body for action in *talk-and-bodily-conduct-in-interaction*. Future research could explore the interactional work that can be accomplished with other components of what is understood as "shrugs," such as mouth shrugs, head shakes, or head tilts. This would enable us to pinpoint the specific functions performed by different prototypical components and to better understand how each of them contributes to "shrugs" as complex ensembles in social interaction.

Data availability statement

The data extracts presented in this study are from the Database of Spoken German, Forschungs- und Lehrkorpus gesprochenes Deutsch (FOLK) and can be found online at dgd.ids-mannheim.de.

Ethics statement

Data from private research corpora was collected after ethical review and approval by Human Research Ethics Review Board at the University of Waterloo (Waterloo, Canada) and Committee on Research Involving Human Subjects at Kansas State University (Kansas, USA). The studies were conducted in accordance with the local legislation and institutional requirements. Ethical review and approval were not required for the use of the publicly available corpus FOLK hosted at the Leibniz Institute for the German Language, in accordance with the local legislation and institutional requirement. All names appearing in the transcripts have been anonymized. Written informed consent for scientific use of the data was obtained from all research participants.

Author contributions

EB: Writing – original draft, Writing – review & editing. AG: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The publication of this article was funded by the Leibniz Institute for the German Language and the Open Access Fund of the Leibniz Association. Emma Betz's work was supported by Canada Foundation for Innovation funding for the creation of a Social Interaction, Language, and Culture Lab at the University of Waterloo (CFI-JELF project #37510).

Acknowledgments

We are grateful to the participants of the conference "Knowledge in talk-in-interaction 2023" in Lausanne, Switzerland, the workshop "Building a Collection" at the 2023 GAL Research School "Interactional research in context of methodological and technical innovation" in Mannheim, Germany, and the 2024 workshop "(Un-)Cooperative Behavior in Social Interaction," also in Mannheim, for their input at earlier stages of the analysis. We thank Galina Gostrer for sharing additional extracts to include in our collection. We thank the two reviewers and our guest editors for their valuable comments on prior versions 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.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

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

Alm, M. (2015). "Also als finale Partikel im Deutschen," in *Das Nachfeld im Deutschen: Theorie und Empirie*, ed. H. Vinckel-Roisin (Berlin, München, Boston: De Gruyter), 319–342.

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). Multimodal stance-taking in interaction–a systematic literature review. *Front. Commun.* 8:1187977. doi: 10.3389/fcomm.2023.1187977

Baldauf-Quilliatre, H., and Imo, W. (2020). "Pff," in *Prosodie und Konstruktionsgrammatik*, eds. W. Imo and J. Lanwer (Berlin; Boston, MA: De Gruyter), 201–232.

Betz, E. (2017). "Diskursmarker aus konversationsanalytischer sicht: prosodisch integriertes *ja* am beginn von responsiven turns," in *Diskursmarker im Deutschen. Reflexionen und Analysen,* eds. H. Blühdorn, A. Deppermann, H. Helmer, and T. Spranz-Fogasy (Göttingen: Verlag für Gesprächsforschung), 183–205.

Betz, E., and Deppermann, A. (2021). "Okay in responding and claiming understanding," in *OKAY Across Languages: Toward a Comparative Approach to its Use in Talk-in-Interaction,* eds. E. Betz, A. Deppermann, L. Mondada, and M.-L. Sorjonen (Benjamins), 55–92.

Cekaite, A. (2010). Shepherding the child: embodied directive sequences in parent–child interactions. *Text Talk* 30, 1–25. doi: 10.1515/text.2010.001

Clayman, S. E. (2010). Address terms in the service of other actions: the case of news interview talk. *Discourse Commun.* 4, 161–183. doi: 10.1177/1750481310364330

Clayman, S. E. (2024). "Working with collections in conversation analysis," in *The Cambridge Handbook of Methods in Conversation Analysis*, eds. J. D. Robinson, R. Clift, K. H. Kendrick, and C. W. Raymond (Cambridge: Cambridge University Press), 191–216.

Clift, R. (2014). Visible deflation: embodiment and emotion in interaction. *Res. Lang. Soc. Interact.* 47, 380–403. doi: 10.1080/08351813.2014.958279

Clift, R. (2021). Embodiment in dissent: the eye roll as an interactional practice. *Res. Lang. Soc. Interact.* 54, 261–276. doi: 10.1080/08351813.2021.1936858

Clift, R. (2023). "Embodied skepticism: facial expression and response relevance," in *New Perspectives on Goffman in Language and Interaction: Body, Participation and the Self,* eds. L. Mondada and A. Peräkylä (New York, NY: Routledge), 170–194

Clift, R. (2024). What is in a look? The accountability of gaze in trajectories to conflict. *Front. Psychol.* 15:1436191. doi: 10.3389/fpsyg.2024.1436191

Corbin, J. M., and Strauss, A. (1990). Grounded theory research: procedures, canons, and evaluative criteria. *Qual. Sociol.* 13, 3–21. doi: 10.1007/BF00988593

Couper-Kuhlen, E. (2009). "A sequential approach to affect: the case of 'disappointment'," in *Talk in Interaction—Comparative Dimensions*, eds. M. Haakana, M. Laakso, and J. Lindström (Helsinki: Suomalaisen Kirjallisuuden Seura), 94–123.

Debras, C. (2017). The shrug: forms and meanings of a compound enactment. *Gesture* 16, 1–34. doi: 10.1075/gest.16.1.01deb

Debras, C. (2021). Multimodal profiles of *je (ne) sais pas* in spoken French. *J. Pragmatics* 182, 42–62. doi: 10.1016/j.pragma.2021.05.018

Debras, C., and Cienki, A. (2012). Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking. *Proceedings of the ASE 2012 International Conference of Social Computing*, Amsterdam, Netherlands, 3–5.

Deppermann, A., and Gubina, A. (2021). When the body belies the words: embodied agency with *darf/kann ich*? ("may/can I?") in German. *Front. Commun.* 6:661800. doi: 10.3389/fcomm.2021.661800

Du Bois, J. W. (2007). "The stance triangle," in *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. R. Englebretson (Benjamins), 139–182. doi: 10.1075/pbns.164.07du

Du Bois, J. W., and Kärkkäinen, E. (2012). Taking a stance on emotion: affect, sequence, and intersubjectivity in dialogic interaction. *Text Talk* 32, 433–451. doi: 10.1515/text-2012-0021

Edwards, D. (2005). Moaning, whinging and laughing: the subjective side of complaints. *Discourse Stud.* 7, 5–29. doi: 10.1177/1461445605048765

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi: 10.1515/semi.1969.1.1.49

Ford, C. E., Thompson, S. A., and Drake, V. (2012). Bodily-visual practices and turn continuation. *Discourse Processes* 49, 192–214. doi: 10.1080/0163853X.2012.654761

Givens, D. (1977). Shoulder shrugging: a densely communicative behavior. *Semiotica* 19, 13–29. doi: 10.1515/semi.1977.19.1-2.13

Glenn, P., and Holt, E. (2013). *Studies of Laughter in Interaction.* London: Bloomsbury.

Golato, A. (2011). "Appreciatory sounds and expressions of embodied pleasure used as compliments," in *Pragmatics of Society,* eds. K. Aijmer and G. Anderson (Berlin; Boston, MA: De Gruyter), 361–392.

Golato, A. (2012). German oh: marking an emotional change of state. *Res. Lang. Soc. Interact*. 45, 245–268. doi: 10.1080/08351813.2012.699253

Golato, A. (2018). "Turn-initial *naja* in German," in *Between Turn and Sequence: Turn-initial Particles Across Languages,* eds. M. L. Sorjonen and J. Heritage (Amsterdam: John Benjamins), 413–444.

Goodwin, M. H., and Alim, H. S. (2010). 'Whatever (neck roll, eye roll, teeth suck)': the situated coproduction of social categories and identities through stancetaking and transmodal stylization. *J. Linguist. Anthropol.* 20, 179–194. doi: 10.1111/j.1548-1395.2010.01056.x

Goodwin, M. H., and Cekaite, A. (2013). Calibration in directive/response sequences in family interaction. *J. Pragmatics* 46, 122–138. doi: 10.1016/j.pragma.2012.07.008.

Groß, A., Dix, C., Ruusuvuori, J., and Peräkylä, A. (2023). Facial gestures in social interaction: introduction to the special issue. *Soc. Interact. Video-Based Stud. Hum. Sociality* 6. doi: 10.7146/si.v6i3.142894

Gubina, A., and Betz, E. (2021). What do newsmark-type responses invite? The response space after German echt. *Res. Lang. Soc. Interact*. 54, 374–396. doi: 10.1080/08351813.2021.1974745

Heller, V. (2018). Embodying epistemic responsibility. The interplay of gaze and stance-taking in children's collaborative reasoning. *Res. Child. Soc. Interact*. 2, 262–285. doi: 10.1558/rcsi.37391

Heller, V., Schönfelder, N., and Robbins, D. (2023). Displaying a critical stance: eyebrow contractions in children's multimodal oppositional actions. *Soc. Interact. Video-Based Stud. Hum. Sociality* 6. doi: 10.7146/si.v6i3.143027

Helmer, H., Betz, E., and Deppermann, A. (2021). "Coordination of okay, nods, and gaze in claiming understanding and closing topics," in *Okay Across Languages: Toward a Comparative Approach to its Use in Talk-in-Interaction,* eds. E. Betz, A. Deppermann, L. Mondada, and M.-L. Sorjonen (Benjamins), 364–393.

Hepburn, A., and Potter, J. (2007). Crying receipts: time, empathy, and institutional practice. *Res. Lang. Soc. Interact*. 40, 89–116. doi: 10.1080/08351810701331299

Heritage, J. (2012). Epistemics in action: action formation and territories of knowledge. *Res. Lang. Soc. Interact.* 45, 1–29. doi: 10.1080/08351813.2012.646684

Hoey, E. M. (2014). Sighing in interaction: somatic, semiotic, and social. *Res. Lang. Soc. Interact.* 47, 175–200. doi: 10.1080/08351813.2014.900229

Hoey, E. M. (2020). Waiting to inhale: on sniffing in conversation. *Res. Lang. Soc. Interact.* 53, 118–139. doi: 10.1080/08351813.2020.1712962

Hömke, P., Holler, J., and Levinson, S. C. (2017). Eye blinking as addressee feedback in face-to-face conversation. *Res. Lang. Soc. Interact.* 50, 54–70. doi: 10.1080/08351813.2017.1262143

Imo, W. (2011). Cognitions are not observable–but their consequences are: Mögliche Aposiopese-Konstruktionen in der gesprochenen Alltagssprache. *Gesprächsforschung–Online-Zeitschrift zur verbalen Interaktion,* 12, 265–300.

Jefferson, G. (2004). "Glossary of transcript symbols with an introduction," in *Conversation Analysis: Studies from the First Generation,* ed. G. H. Lerner (Amsterdam; Philadelphia, PA: Benjamins), 13–31.

Jefferson, G., Sacks, H., and Schegloff, E. (1977). *Preliminary notes on the sequential organization of laughter. (Pragmatics Microfiche)*. Cambridge: Cambridge University, Department of Linguistics.

Jehoul, A., Brône, G., and Feyaerts, K. (2017). The shrug as marker of obviousness: corpus evidence from Dutch face-to-face conversations. *Linguist. Vanguard* 3:20160082. doi: 10.1515/lingvan-2016-0082

Kaukomaa, T., Peräkyl,ä, A., and Ruusuvuori, J. (2013). Turn-opening smiles: facial expression constructing emotional transition in conversation. *J. Pragmatics* 55, 21–42. doi: 10.1016/j.pragma.2013.05.006

Kaukomaa, T., Peräkylä, A., and Ruusuvuori, J. (2014). Foreshadowing a problem: turn-opening frowns in conversation. *J. Pragmatics* 71, 132–147. doi: 10.1016/j.pragma.2014.08.002

Keevallik, L., and Ogden, R. (2020). Sounds on the margins of language at the heart of interaction. *Res. Lang. Soc. Interact.* 53, 1–18. doi: 10.1080/08351813.2020.1712961

Kendon, A. (2002). Some uses of the head shake. *Gesture* 2, 147–182. doi: 10.1075/gest.2.2.03ken

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Marrese, O. H., Raymond, C. W., Fox, B. A., Ford, C. E., and Pielke, M. (2021). The grammar of obviousness: the palm-up gesture in argument sequences. *Front. Commun.* 6:663067. doi: 10.3389/fcomm.2021.663067

Mondada, L. (2018). Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51, 85–106. doi: 10.1080/08351813.2018.1413878

Mondada, L. (2024). "Appendix II: multimodal transcription conventions," in *The Cambridge Handbook of Methods in Conversation Analysis*, eds. J. D. Robinson, R. Clift, K. H. Kendrick, and C. W. Raymond (Cambridge University Press), 990–998.

Mondada, L., and Sorjonen, M.-L. (2021). "OKAY in closings and transitions," in *OKAY Across Languages: Toward a Comparative Approach to its Use in Talk-in-interaction,* eds. E. Betz, A. Deppermann, L. Mondada, and M.-L. Sorjonen (Amsterdam; Philadelphia, PA: Benjamins), 94–127. doi: 10.1075/slsi.34.04mon

Morris, D. (1994). *Bodytalk: A World Guide to Gesture*. London: Jonathan Cape.

Müller, C. (2004). "Forms and uses of the palm up open hand: a case of a gesture family?" in *The Semantics and Pragmatics of Everyday Gestures*, eds. C. Müller and R. Posner (Berlin: Weidler), 233–256.

Ogden, R. (2013). Clicks and percussives in English conversation. *J. Int. Phon. Assoc.* 43, 299–320. doi: 10.1017/S0025100313000224

Ogden, R. (2020). Audibly not saying something with clicks. *Res. Lang. Soc. Interact.* 53, 66–89. doi: 10.1080/08351813.2020.1712960

Pekarek Doehler, S. (2022). Multimodal action formats for managing preference: *chais pas* 'dunno' plus gaze conduct in dispreferred responses to questions. *J. Pragmatics* 197, 81–99. doi: 10.1016/j.pragma.2022.05.010

Pittner, K. (2006). "Common ground in interaction: the functions of medial *doch*," in *Lexical Markers of Common Ground,* eds. A. Fetzer and K. Fischer (Oxford; Amsterdam: Elsevier), 67–87.

Pomerantz, A. (1986). Extreme case formulations: a way of legitimizing claims. *Hum. Stud.* 9, 219–229. doi: 10.1007/BF00148128

Raymond, C. W., and Robinson, J. D. (2024). "Evidencing conversation-analytic claims: how participants orient to social action," in *The Cambridge Handbook of Methods in Conversation Analysis*, eds. J. D. Robinson, R. Clift, K. H. Kendrick, and C. W. Raymond (Cambridge: Cambridge University Press), 315–355.

Raymond, G. (2004). Prompting action: The stand-alone" so" in ordinary conversation. *Res. Lang. Soc. Interac.* 37, 185–218. doi: 10.1207/s15327973rlsi3702_4

Reber, E. (2012). *Affectivity in Interaction: Sound Objects in English*. Amsterdam; Philadelphia, PA: Benjamins. doi: 10.1075/pbns.215

Reber, E., and Couper-Kuhlen, E. (2020). On "whistle" sound objects in English everyday conversation. *Res. Lang. Soc. Interact.* 53, 164–187. doi: 10.1080/08351813.2020.1712966

Robinson, J. D. (2016). *Accountability in Social Interaction*. Oxford: Oxford University Press.

Robinson, J. D., Clift, R., Kendrick, K. H., and Raymond, C. W. (eds.). (2024). *The Cambridge Handbook of Methods in Conversation Analysis.* Cambridge: Cambridge University Press.

Sacks, H. (1984). "On doing "being ordinary"," in *Structures of Social Action: Studies in Conversation Analysis,* eds. J. M. Atkinson and J. Heritage (Cambridge University Press), 413–429.

Sacks, H. (1992). *Lectures on Conversation* (Vol. 1 and 2). Oxford: Basil Blackwell.

Schegloff, E. A. (1996a). "Turn organization: one intersection of grammar and interaction," in *Interaction and Grammar*, eds. E. Ochs, E. A. Schegloff, and S. Thompson (Cambridge University Press), 52–133.

Schegloff, E. A. (1996b). Confirming allusions: toward an empirical account of action. *Am. J. Sociol.* 102, 161–216. doi: 10.1086/230911

Schegloff, E. A. (2007). *Sequence Organization in Interaction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511791208

Schmidt, T. (2016). Good practices in the compilation of FOLK, the research and teaching corpus of spoken German. *Int. J. Corpus Linguist.* 21, 396–418. doi: 10.1075/ijcl.21.3.05sch

Schoonjans, S. (2018). *Modalpartikeln als Multimodale Konstruktionen: Eine Korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen*. Berlin; Boston, MA: De Gruyter.

Sidnell, J., and Stivers, T. (eds.). (2012). *The Handbook of Conversation Analysis.* Oxford: Wiley-Blackwell.

Skogmyr Marian, K. (2021). Assessing without words: verbally incomplete utterances in complaints. *Front. Psychol.* 12:689443. doi: 10.3389/fpsyg.2021.689443

Stevanovic, M., and Peräkylä, A. (2014). Three orders in the organization of human action: on the interface between knowledge, power, and emotion in interaction and social relations. *Lang. Soc.* 43, 185–207. doi: 10.1017/S0047404514000037

Stivers, T. (2022). *The Book of Answers: Alignment, Autonomy, and Affiliation in Social Interaction*. New York, NY: Oxford University Press.

Stivers, T., and Rossano, F. (2010). Mobilizing response. *Res. Lang. Soc. Interact.* 43, 3–31. doi: 10.1080/08351810903471258

Streeck, J. (2009). *Gesturecraft: The Manu-facture of Meaning*. Amsterdam; Philadelphia, PA: Benjamins.

Tennant, R. A., and Brown, M. G. (1998). *The American Sign Language Handshape Dictionary*. Washington, DC: Gallaudet University Press.

Wiggins, S. (2002). Talking with your mouth full: gustatory *mmm*s and the embodiment of pleasure. *Res. Lang. Soc. Interact.* 35, 311–336. doi: 10.1207/S15327973RLSI3503_3

Wiggins, S. (2013). The social life of 'eugh': disgust as assessment in family mealtimes. *Br. J. Soc. Psychol.* 52, 489–509. doi: 10.1111/j.2044-8309.2012.02106.x

Wilkinson, S., and Kitzinger, C. (2006). Surprise as an interactional achievement: reaction tokens in conversation. *Soc. Psychol. Q.* 69, 150–182. doi: 10.1177/019027250606900203

Wright, M. (2011). On clicks in English talk-in-interaction. *J. Int. Phon. Assoc.* 41, 207–229. doi: 10.1017/S0025100311000144

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 11 July 2025
DOI 10.3389/fpsyg.2025.1452460

Check for updates

OPEN ACCESS

EDITED BY
Silva H. Ladewig,
University of Göttingen, Germany

REVIEWED BY
Himmbler Olivares,
University of Concepcion, Chile
Lifang Wei,
Shaoxing University, China

*CORRESPONDENCE
Stefan Pfänder
✉ stefan.pfaender@romanistik.uni-freiburg.de

RECEIVED 20 June 2024
ACCEPTED 28 April 2025
PUBLISHED 11 July 2025

CITATION
Pfänder S and Pfänder C (2025) Embodying togetherness while taking divergent stances. Romantic couples' multimodal positioning practices while performing "we-stories". *Front. Psychol.* 16:1452460. doi: 10.3389/fpsyg.2025.1452460

COPYRIGHT
© 2025 Pfänder and Pfänder. 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.

Embodying togetherness while taking divergent stances. Romantic couples' multimodal positioning practices while performing "we-stories"

Stefan Pfänder* and Caroline Pfänder

Department of Romance Studies, University of Freiburg, Freiburg, Germany

Making epistemic and/or affective statements about an interlocutor is a rather delicate endeavor. This is all the more true for spouses who collaboratively tell a good friend a "we-story" about where they met, when they fell in love, how he proposed to her, and that they were not always good partners in everyday life. Using a corpus of 48 collaborative narratives of Italian romantic couples' we-stories, we examine how strong epistemic and affective standpoints interrupt the narrative flow and open up a side sequence in which the delicate positioning of the other is multimodally constructed and negotiated. Using multimodal conversational analysis of three exemplary excerpts, we show how the possibilities of sitting side by side on a sofa while recounting difficult marital episodes affect the interplay of verbal, vocal, and bodily resources in the conversational interaction. Faced with a potentially face-threatening act, participants make use of remarkable multimodal packages to challenge their spouse's unwelcome stance-taking by formulating a counter-stance. These opposing stance-takings then lead to a negotiation and ultimately to a new collaborative narrative that most of the times integrates parts of both (initially divergent) stances. We conclude that a finely nuanced micro-sequential analysis makes it possible to discover the highly complex interplay of multimodal resources like verbal and gestural resonance, mutual nodding, synchronized position shifts, eye contact, choral vocalizations and, maybe most importantly, joint laughter. By reusing, but slightly transforming, these verbal and nonverbal elements from prior talk, romantic partners co-operatively achieve shared epistemic and/or affective stance-taking in collaborative story-telling.

KEYWORDS

stance-taking, affective stance, epistemic stance, conversation analysis, embodied practices

1 Introduction

In the embodied practice of jointly telling a friend about where they met, when they fell in love, how he proposed, and about what they quarrel, romantic couples face a severe challenge in talk in interaction. What they are about to tell has been labeled a "we-story" (Gildersleeve et al., 2017; Huber, 2015; Singer and Skerrett, 2014; Strong et al., 2014). We-stories are not easy to tell as they clearly make the stance-taking of both participants relevant, yet, only one person can speak at a time.

One possible solution to this problem that we encounter in our data is that, while one person is telling one of the above-mentioned episodes, their partner employs bodily resources (such as raising their eyebrows, smirking, inhaling deeply, or freezing their upper body, opening their eyes wide etc.) in order to take affective and/or epistemic stances toward the emerging utterance. More often than not, these ephemeral positionings are treated by the current speaker as foreshadowing trouble, or even as conversational challenges. In these cases, the story-telling is momentarily broken off for a side sequence that allows the romantic partners to quickly negotiate how they remember how things took place, and how both of them experienced the narrated event back then, or how they feel about the other person's prior statement in the process of collaborative story-telling.

While taking divergent stances in these side sequences, the partners (in our sample, mostly spouses) not only try to quickly come to a shared understanding of what actually happened, but simultaneously strive at reestablishing both a told and performative stance as a harmonious couple. The maybe most striking result is that, put very simply, while telling you that I don't like what you say, I can bodily show both you and them—i.e., the attentively listening third person as well as imagined later recipients—that we are still in a happy relationship. In a nutshell, then, multimodal stance-taking simultaneously allows for my individual voice to be heard and, at the same time, togetherness to be embodied, such as to make the potentially upcoming anger disappear as fast as possible.

The remainder of this paper is organized as follows: in § 2 we briefly summarize findings on stance-taking in conversation and story-telling. In § 3 we present our data (Italian romantic couples' collaborative story-tellings) and methodology. § 4 explores the semiotic resources used to accomplish an agreement after challenging and negotiating a stance during collaborative story-telling in three exemplary "we-stories". In § 5 we give an overview of our empirical results and, in conclusion, discuss the implications of our observations for a multimodal conception of stance-taking.

2 Stance-taking in conversation and story-telling

Stance-taking is "the public act of positioning oneself toward objects, people or states of affairs" (Andries et al., 2023, p. 1). However, this cannot be done without taking into consideration who we are talking to and what their stance is. Therefore, in his stance triangle, Du Bois (2007) defines the process of stance-taking as "a public act by a social actor, achieved dialogically through communicative means, of simultaneously evaluating objects, positioning subjects (self and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field" (p. 163). In interactional research, we generally differentiate three types of stance: (1) epistemic stance, (2) affective stance, and (3) deontic stance. Epistemic stance is concerned with our knowledge. Therefore, the questions of primary concern are how knowing we are and how knowing we present ourselves to our co-participants (Couper-Kuhlen and Selting, 2018). However, importantly, our epistemic stance is not fixed, but rather, it is a "thoroughly interactional and emergent process" (Couper-Kuhlen and Selting, 2018, p. 4). Affective or emotional stance, on the other hand, is mainly concerned with how we feel toward an object of stance, our emotions and attitudes toward it (Couper-Kuhlen and Selting, 2018). When being expressed in conversation, these emotions are no longer a personal matter, but they become interactionally relevant through their public display. Lastly, deontic stance is concerned with how desirable an action is.

As has already become evident, stance-taking in interaction is always collaborative, as we usually take into account how our co-participants position themselves toward the stance object (Du Bois, 2007). Therefore, as noted by Kärkkäinen (2006), "stance is very often established and negotiated as an interactional practice" (p. 718; cf. also Bröker and Zima, 2022). The stance we take toward a stance object is thus not fixed, but may change, based on the conditions under which the conversation takes place. For instance, in the case of epistemic stance, it might be important for the participants in conversation to establish who knows what, e.g., if we do not know whether our recipient knows more or less on a specific matter than we do (Satti, 2023a). Interestingly, stance-taking can also be requested by one of the participants, e.g., in a request for verification, where the teller of a story asks his co-teller to verify a specific element of the story delivered to a third party (Hügel, 2012; Satti, 2023b). Through this, tellers can request their co-teller to take a stance on a specific element of the story they are telling, which can either reinforce or challenge the current teller's stance.

Our research will show that when taking a stance in we-stories, co-participants are actually very attentive about what their spouse claims about their shared experiences and attitudes as a couple.

When taking a stance, speakers oftentimes draw on lexico-grammatical resources. Such resources include specific phrases such as "I guess" (Kärkkäinen, 2007) or grammatical markers such as modal verbs (Biber and Finegan, 1989). However, we can express a stance not only by what we say, but also by when we say it. For instance, by chiming into the turn of the person we are talking with, we can express that we share their affective stance, e.g., toward a scenario. Interestingly, in this case, we not only show that we share their affective stance, but we also publicly display that we are equally knowing, which shows that epistemic and affective stances can come hand-in-hand in conversation (Pfänder and Couper-Kuhlen, 2019). However, stance-taking is not a purely verbal accomplishment, but rather, different kinds of multimodal resources can also be used to express one's stance.

What is more, the literature on stance-taking suggests that verbal means of stance- taking rarely make up for a stance act on their own, but rather, they are frequently accompanied by multimodal resources (Andries et al., 2023). Such resources include, for example, prosody (Freeman, 2019), gestures (Yang and Wang, 2025), body movements (Trujillo and Holler, 2021), facial expressions (Ruusuvuori and Peräkylä, 2009), or gaze behavior (Haddington, 2006).

It is, however, interesting to note that verbal and multimodal resources do not always express the same stance, but rather, the body can express a very different stance from what our words

do (e.g., Deppermann and Gubina, 2021). This can be done simultaneously, as Andries et al. (2023) have shown convincingly: "The possibility to use multiple semiotic resources (bodily-visual or other) simultaneously, gives rise to a wide range of options for participants to time their stance display, and continuously adapt their stance to that of their interlocutor, without interrupting talk" (Andries et al., 2023, p. 5). In our data of we-stories we more often than not find instances of quasi-simultaneous stance-taking, whereby one of the two only takes a position in terms of body language, which can range from agreement (e.g., nodding, smiling, looking at each other) to astonishment (e.g., putting one's head back, freezing, frowning) to rejection of or dissatisfaction with what is being said.

3 Data and methodological procedure

The *Sofa Talks* Corpus (University of Freiburg) comprises 298 video recordings ranging in duration from 10 to 40 min. The corpus data was extracted from as naturalistic a setting as possible, but within an experimental framework. Participants were invited to sit comfortably on a sofa in the presence of a third, well-known friend or relative, and narrate shared experiences. This allowed for a relaxed and familiar environment in which participants could talk comfortably. However, it is acknowledged that collaboratively telling a story in front of a camera, does not replicate everyday life.

Each video contains two participants in a close relationship, be that siblings, friends, or married couples. Both participants were clearly invited to reminisce and tell stories of shared experiences, thus giving both of them equal status in the conversation and equal right to speak. Only if a truly shared experience was discussed, was the video included in the corpus. As a result, the corpus contains 298 video recordings from conversations in German, French, Spanish, Catalan, and Italian. The current study focuses on the Italian data.

In recent years, the expressive bodily resources that contribute to the emergent design of turns and sequences have been taken more and more seriously. This growing body of research is bringing to the open as more and more cases where certain bodily movements recurrently co-occur with verbal expressions in the design of situated courses of action. This systematic interplay of the verbal dimension and embodied elements has been conceptualized as "multimodal packages" (Goodwin, 2007; Pekarek Doehler, 2019; Hofstetter and Keevallik, 2020; Stevanovic, 2021; other authors use the concept of multimodal gestalts; cf. Mondada, 2015; Stukenbrock, 2021). Following Stevanovic (2021, p. 2), we sustain that "an essential feature of such multimodal formations is that none of their single components can achieve the given action on its own. In many cases, these formations become conventional practices for achieving certain goals within a community or activity".

Some highly conventional practices are recognizable through emblematic gestures,[1] whereas others (as those under scrutiny here), are more context sensitive, but still recurrent (Satti, 2023a; Ladewig, 2014, 2024).

The crucial importance of the moment-by-moment unfolding of emergent utterance in the real time of interaction has been proven over and over in Conversation Analysis and Interactional Linguistics (Couper-Kuhlen and Selting, 2018). More recently, it has become clear that the bodily expressive movements are equally sensitive to the dialogic temporality (Deppermann and Streeck, 2018). However, the temporality of bodily expressions differs in at least two ways from the temporal design of verbal utterances. First, bodily movements seem to be less bound to turn constructional units. They more often than not start before the verbal utterance and can continue or fade out afterwards. Bodily expression thus can both project or foreshadow verbal expressions that are about to come (Kaukomaa et al., 2014) and they can frame them after the utterance (cf. Pfänder, 2023). Second, the communicative function of bodily movements can be attributed to at least four different dimensions, namely intercorporeality, coordination, common ground, and co-semiosis (for a similar account of interactional dimensions, cf. Pfänder, 2023; and Meyer, 2014). Collaborative story-telling under scrutiny here extensively relies on expressive resources for displaying *intercorporeality*, a shared embodied experience (Tanaka, 2016) of the co-operative actions that unites speakers as they engage (Goodwin, 2018). This requires subtle dynamics of more often than not kinesically achieved (micro-sequential) *coordination*, ensuring that participants know when there is a good moment to take a turn (Deppermann and Schmidt, 2021). Successful interaction can unfold only if *common ground* is constantly being established (Clark and Brennan, 1991; Clark, 1996/2012), meaning that all participants share a similar understanding of what their counterpart is talking about and what they intend to convey. And last, not least, these collaborative story-tellings live by *co-semiosis*, i.e., the collaborative effort to develop the topic of conversation and to create sense together (Schmid, 2020).

For our study, we have made a collection of 48 instances of multimodal stance-taking, of which we discuss—by way of exemplification—three instances in the following Section 4. These excerpts have been chosen to exemplify the three most common types of sequential outcome formats of negotiating stances in our romantic couples' collaborative story-telling sample, namely (a) retracting the counter-stance and agreeing on the initial stance, (b) achieving a new shared stance that integrates parts of both the initial stance and the counter-stance, or (c) slightly changing the topic under discussion.

4 Multimodal stance-taking in we-stories

Singer and Skerrett (2014) and Gildersleeve et al. (2017) worked out a concise definition of we-stories: "A We-Story is a type of couple narrative composed by both partners that describes a vivid shared memory. These stories often provide an important image, metaphor, or phrase that serves as a touchstone for the relationship, and they embody the love and commitment each partner feels for the other" (Gildersleeve et al., 2017, p. 314).

1 Stevanovic (2021, p. 2) gives the following examples: "the striking of the hammer to conclude an auction sale or to initiate a move to a next item in the meeting agenda completing a turn-at-play by placing a token on the game board or formalizing decisions by writing them down" (references omitted).

Consider, for instance, excerpt 1. In the chronological narrative from meeting to marriage, told by the wife on behalf of both of them, she pauses briefly, and then tells how they started to plan their future together, without being engaged, at least formally. Elena utters her epistemic stance: she does not remember that Paolo really proposed marriage. This is where the excerpt[2] starts:

2 All three excerpts are transcribed following the GAT2-system, Selting et al. (2009); and Mondada (2018) for the multimodal resources.

Excerpt 1: ***Marriage proposal***

ELE = Elena, PAO = Paolo

01 ELE: [non] è che c'è stata proprio una proPOSta-=
it's not like there's actually been a proposal
02 PAO: [sì.]
yes
03 ELE: =da +por parte di qualCUn[+o:: ci] spoSIAmo; °h
from anyone "let's marry"
ele: +pointing gesture @PAO+
04 SAR: [((laughs))]
05 PAO: ((smiles broadly @SAR))

Elena accompanies her statement "it's not like there's actually been a proposal from anyone" [non] è che c'è stata proprio una proPOSta da (.) por parte di qualCUno with a glance at her husband, who reacts with a questioning look. While eye contact is being made, the third person (SAR), a friend of both, laughs briefly. As a reaction to the laughter Paolo casts a broad smiling glance at SAR. In the meanwhile, Elena then expands her statement and, by animating an imaginary figure, makes it clearer what she meant by her statement, i.e., she gives an account. She has an imagined man (who was not her husband) say quietly: "let's marry" ci spoSIAmo (l.03). This is what in her imagination would have been an "actual marriage proposal". But instead she remembers that it was "something a bit" … and she searches for the right adjective to complete the sentence (l.06); instead she completes her utterance multimodally, waving her hands and leaving them in the air, as if things were not spelled out clearly, rather the message was somehow "in the air":

06 ELE: è stato(.)+una cOsa un po[::-°+]
it was kind of like
ele: + palm-up in the air+

In dialogic resonance, i.e., using the same syntactic construction "let's V O", Paolo animates himself back then and remembers what he said at the time: "Let's buy a house" compriamo CAsa (l.07). His wife confirms this with a smile (l.08).

Paolo continues and evaluates his way of proposing marriage with the assessment: "well that was obvious" be era OVvio (l. 11). To accompany his speech, the speaker performs a "cycle gesture" with both hands (fig. 01), which ends in an "obvious gesture" (cf. Marrese et al., 2021, i.e., "both hands palm up, on hold" as can be appreciated in fig. 02):

07 PAO: [co]mpriamo CAsa,
"let's buy a house"
08 ELE ((smiles))
09 PAO: mhm-
uhm
10 (1.0)
11 PAO: be era %°OVvi[o. °]
well it was obvious
pao: %gesture–>
fig: °fig.02 °fig.03

Elena laughs and repeats—each time with an adaptation—both the verbal construction and the gestural dynamic. She utters—now turning back to the third person— (especially in the Italian original) a syntactically very similar but semantically different construction "You could guess it" <<:-)>era da intuIre.> (l. 12). Just like the verbal construction, the gesture begins in a very similar way with a cycle gesture (compare fig. 02 and fig. 04), which, however, does not end in an "it's obvious" stroke as his gesture, but dissolves into a much larger gesture on both sides, and which fits the statement "You could guess it":

12 ELE: **[<<:-)> %e]ra +da intuI°[re.>** **]**
you could guess it
(lit: it was to be guessed)
ele: +gesture-->
fed: -->%
fig: °fig.04

13 SAR: **[((laughs))]**
[((laughs))]
14 ELE: **[((laughs))]**+
ele: -->+

A movement of mediation takes place, which works conciliatorily in two directions: first to the partner, then to the camera, i.e., to the public, in the sense of a rehabilitation measure for the partner, which ensures that Paolo does not suffer a loss of reputation in the eyes of the public. Both laugh heartily at this conciliatory moment that ends the side sequence of negotiation with Elena's stance taking in line 15: "It was said between the lines": era dEtto tra le RIghe.

15 era dEtto tra le RIghe;
it was said between the lines
16 °hh
17 eh:::-
umm:
18 <<all> e quindi insomma NIENte;>=
and then well nothing
19 =ci siamo messi a cercAre un po CAse::-
we started looking a bit for houses

In a nutshell then, the negotiation of stances is done by a variety of instances of verbal and bodily resonance (Brône and Zima, 2014), the opponents repeat and thus reuse the same multimodal resources but change the course of action by coming to a different end and thus expressing a different stance. Since Paolo wanted to buy a house with her, it was clear to him that there was no need for an explicit marriage proposal. Elena, on the other hand, emphasizes that there was no explicit proposal and that her common sense was needed to understand the house purchase as an expression of the desire to enter into a life bond with her, which was only revealed to her between the lines.

Thus, it is only through a multimodal analysis of stance and counter stance (epistemic and affective-evaluative) that we see that this is not a disruptive moment in a couple telling their love story, but rather a humorous form by integrating his cycling gesture with her uncertainty movement in one complex gesture trajectory. As we advocate the view that language is inherently multimodal and thus consider utterances to be of composite nature (Enfield, 2013), verbal and bodily expressive resources are, with Enfield's words, "draw[n] [...] together into unified, meaningful packages" (2013, p. 689). The multimodal negotiation of stance-takings ensures that the we-story actually remains a shared story in which the love and commitment for the spouse is expressed. Overall, it can be said that the positioning is found in the verbal wording, the gestures and body movements express the journey from stance to counter stance and, ultimately from separation via negotiation to reconnection.

The next example, then, again shows how couples can agree to disagree, but in a slightly different way. In this example, Valentina and Manolo are talking about how often they met during the time Valentina studied in Milan.

Excerpt 2: *Four times*

VAL = Valentina, Man = Manolo[3]

3 Sofa-Talks Ita_2017_Dantoni_02 7.09 – 7.32.

112 MAN: en tre anni que sono stato giù?
in three years I came to see you (only three times)
113 VAL: ((laughs))
114 MAN: %eh NO.
no way
man: %grappolo gesture–->
115 VAL: ((laughs))%
man: –->%
116 MAN: che bugiarda
what a liar (you are)
117 VAL: bon più di TRE o quattro volte non sEi venuto giù
well you didn't come to see me more than 3 or 4 times
[manolo.]
, manolo
118 MAN: [ma sei] FUOri;
you're out of your mind
119 VAL: ((coughs))

Having reconstructed that they have been together for 5 years, Valentina and Manolo jointly remember the early years when they used to study at different universities and had to travel several hours by car to see each other.

Valentina then makes the affective, emotional, and somehow deontic-evaluative stance that what had made her insecure about the future of their relationship was that Manolo only came to see her three or four times during the first 2 years. Manolo disconfirms this claim and insists that he came far more often. She then asks him to specify how often he actually came to see her and he gives the unprecise answer of "often enough". She insists that it had been no more than three or four times, showing this number by tipping her fingers. The use of "only" (solo) has the effect of threatening his face. In order to deal with this injury, he first adopts a dismissive attitude (eh NO), which is then intensified in such a way that he describes his partner as a liar (che bugiarda) and then as someone who is FUOri, i.e., "out of their mind". Finally, he follows up on her request to know how often he did visit her back then and what he was initially unable to remember (non mi riCORdo), he now can (mi ricordo). She laughs knowingly and challenges him to neatly reconstruct the times he actually came to visit her.

120 QUANte allora dimmi [tU-]
How many times? You tell me then.
121 MAN: [eh] non mi riCORdo;
I don't remember
122 VAL: ((coughs))
123 ((laughs))
124 sil: (--)
125 MAN: mi ricordo- [(0.3)]
I remember
VAL: [((coughs))]

He takes up her way of illustrating each visit by touching one of the fingers of his hand, giving details about each visit: once it rained, another time there was fog in the street, yet another time was in spring and the fourth time at the end of summer.

126 MAN: allora %una vOlta pioVEva%-=°
one time it was raining
man: %first finger list%
fig: °fig.05

127 =%una [vOlta: c'era la NEB<<:-)>bia%°,>]
[((laughs))]
one time it was foggy
man: %second finger list----------%
128 VAL: [((risa))]
fig: °fig.06

129 [ah bon] SÌ.
ah well yeah
130 MAN: %pOi periodi: di primaVEra%-°
then in spring times
man: %third finger list--------%
fig: °fig.07

131 %e: fine estate %°
and at the end of summer
man: %fourth finger list%
fig: °fig.08

132 sil: (1.6)

Then she mocks his imitation of her pointing at each finger, counting from one to four and summing it up by a quick gesture covering all her fingers with the other hand, and uttering that this was not really oftentimes.

133 VAL: si ma °questo una VOLta;(.)
yes but that (means) one time
fig: °fig.09

She restarts by saying it was four times and he repairs her statement saying it was six times. They both laugh out loud and start to narratively reconstruct the first week, agreeing that this was a wonderful shared experience. Again, here, we have a long negotiation resolved in an agreement to disagree about how many times he visited her, closing this sequence and opening another sequence of jointly reconstructing the first weekend they spent together. The multimodal character of the negotiation is at first a means of making the epistemic stance literally more concrete, the gesture resonance is then used as a means of mocking via imitation. Thus, carrying out the same gestures leads them to different conclusions which are verbally uttered.

For Valentina this is a complete list of the visits that actually happened; for Manolo, it is only what he remembers, but there were certainly more visits. Posture shows this, he leans back engulfed in the memories while she leans forwards showing him the facts. That is why this sequence ends with 4 vs. 6.

134 una per [VOLta non è <<f>che::>-]
one time per season is not (much)
135 MAN: [per CUI bon va] <<ff>bEh::> adEs[so::-]
that's why, well, come on now
136 VAL: [quattro]
VOLte.
four times
137 MAN: !SEI!.
six
138 VAL: comunque.
whatever
139 ((laughter))

In a nutshell remember their experience differently which leads them to utter divergent epistemic stances: While Valentina remembers 4 visits, so it was 4 in total, Paolo remembers 4 so there must have been more. She evaluates the number as insufficient. They both count to 4, but for her, it's about the total number, he remembers individual episodes, he remembers the weather, the seasons, etc. By doing so, he protects himself against a potential accusation of misbehavior underlying his wife's stance-taking.

There is one main difference between this example and the previous one about the marriage proposal. Specifically, in this example, we have some movement but very different sense making, while in the previous example, we had movements and utterances, starting out alike but changing on the fly. Then again both examples are very similar in that the multimodality allows the dealing with opposite stances in a humorous way, resulting each time in an agreement to disagree.

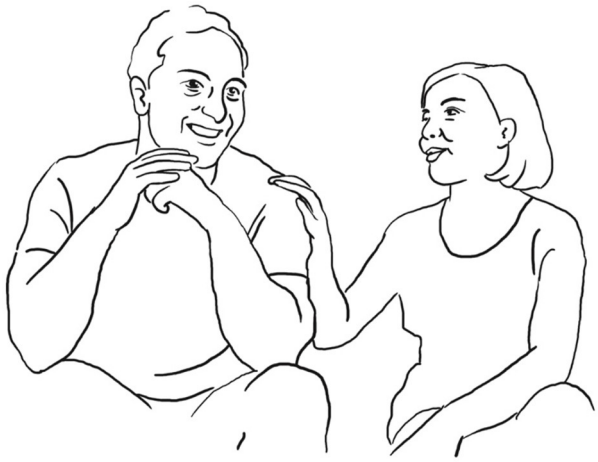
In the next example, the situation is slightly different. Here, Angelina and Luigi tell about their time as a newly wedded couple.

Excerpt 3: ***She breaks my balls, . . . but only when she is tired***
ANG = Angelina, LUI = Luigi[4]

Angelina begins to tell a story making fun of him just as he made fun of her back then, 2 days after their wedding. Telling this we-story to a third person, her friend, also an Italian woman living in Germany, she starts the stance-taking sequence, polyphonically reenacting his words.

4 Sofa-Talks Ita_2017_Dantoni_04 chi lava i piatti.

39 ANG: =<<:-)> %mi ha DETto,>°
he told me
ang: %puts hand on LUI's shoulder-->
fig: °fig.10



40 °h
41 <<acc>due GIORni dopo che->
two days after
42 NO:;
no
43 un po' di PIÙ-=
a bit early
44 =<<acc> che ci siamo spoSAti->%
after we got married
ang: -->%
45 sicCOme abbiamo fatto diecimila cose no,
as we had been doing ten thousand things right
46 SAR: (-)
47 ANG: %e [poi mi] FA,
and then he says to me
ang: %puts hand on LUI's shoulder->
48 LUI: [NO:]
no
49 che il-
that the
50 ANG: SENti-
Listen

She literally gets in contact, touching him on the shoulder (fig. 10), doing the good friends gesture, and searching for eye-contact, so that togetherness is established as a solid basis for the now starting, possibly face-threatening reenactment of a conflict encounter only 2 days after their wedding. She explains why she was a bit annoying during the course of the wedding preparations:

51 LUI: $**NO-°** $
No
lui: $touches her hand$
fig: °fig.11

52 ANG : <<a>sE il: tuo% GRAdo?>
if your degree
ang: -->%gesture thumb + index finger-->
53 LUI: °h
54 ANG: di diventAre rompiPALle:-°
of becoming a pain in the neck
fig: °fig.12

55 SAR: (-)
56 LUI: <<:-)> he->%
ang: ->%
57 ANG: CREsce-=
grows
58 LUI: [in maNIEra esponenziale-]
in an exponential manner
59 ANG: [=così TANto,]
so much

Luigi then downgrades her evaluating stance by making an account for her getting annoying, stating that she is only difficult when tired quando è STANca (l. 69). She acknowledges and they had to organize so many things, in her words "ten thousand things" diecimila cose (l. 45), that she possibly got nervous.

He tries several times to establish himself as a co-teller of the sequence and finally manages to take the turn by uttering "no" NO: (l. 48) and at the same time touching her hand on his shoulder (fig. 11).

makes the "precision" gesture (fig. 12). He puts an end to this possibly face-threatening episode uttering NO-, brushing away his wife's gesture (l. 60, fig. 13, cf. Bressem and Müller, 2014):

60 LUI: $[NO-]° $
lui: $brushes her hand away$
no
fig: °fig.13

61 ANG: [es]ponen[ZIAle;]
Exponentially

From here, Luigi steps out of the story-telling activity, and goes on describing his partner as una ragazza fanTAStica. This verbal compliment alone is strong enough to express his positive relationship with his wife, no additional touching or physical expression is required, and so he succeeds in overwriting his words that hurt his partner 2 days after the wedding as well as his own injury in the interview (he is presented in a bad light due to his partner's story) with his statement and creating a harmonious atmosphere between the couple, which the public should also experience.

62 LUI: [la veri]tà è che:-
the truth is that
63 ANG: ((risa))
64 LUI: ((click))
65 ((click))
66 è una ragazza fanTAStica;
she is a fantastic woman
67 ANG: oh (.) GRAzi[e; ((risa))]
oh thank you
68 LUI: [peRÒ:-]
but
69 quando è STANca-
when she is tired
70 SAR: (-)
71 ANG: ROMpo i co (.) eh.
I'm a pain in the ne(ck)
72 LUI: [$diventa un po' rompi coGLIOni;]°
she becomes a bit of a pain in the neck
lui: $gesture open hands-->
fig: °fig.14

73 ANG: [((laughing))]

74 LUI: [<<laughing> è STA (.)>]
and is
75 ANG: [((laughter))]
76 LUI: [solo quando è STANca; $]
only when she is tired
lui: -->$
77 ANG: ((laughter)) °hh
78 LUI: se nO è una raGAZza °h-
if not she is someone
79 [una raGAZza: sì]:
someone yes
80 ANG: [tranQUILla dai;]
calm right
81 LUI: [tran]QUIL[la (sì) s-]
calm yes
82 ANG: [°h]

There is a lot of co-construction taking place throughout the whole micro-sequence. The husband succeeds in taking part in the joint stance-taking action by putting his hand on her hand on his shoulder and uttering "no" several times. Both showing a wide smile. When he has finally won the right to speak, he changes the gaze direction as if to not share the story with his wife, but with his wife's friend sitting opposite them. He avoids gazing at his wife and somehow re-writes their we-story alone.

The plethora of multimodal resources employed by the participants has at least two communicative functions, one is to explicitly establish bodily contact through mutual touch, synchronized wide smiles and eye contact, and second, to negotiate the right to speak and finally to co-construct step by step a shared version of their evaluative stance-taking of her in his eyes, both laughing as he looks into the camera and she in his direction (fig. 14). The observer notes a clear release in the bodily tension as if both participants were happy to have overcome the delicate moment.

As she recounts the story that casts him in a negative light, she touches his shoulder, as if to show: "But I love him anyways". He touches her back by placing his hand on hers, simultaneously uttering NO:; (l. 42), but does not immediately gain the right to speak. He then completes her emerging sentences, as if to affiliate with her epistemic stance, using this as a means to narrate his own version of the story. Subsequently, she does the same, completing his sentences as if to say: "I, too, know what happened back then". He overrides her statement about him threatening that, if she continued in that manner… (the actual threat remains unspoken). Naturally, he does not want to be attributed with such a remark 2 days after the wedding, especially on camera. He exits the narrative episode and repositions himself within a general "whenever" structure. She can indeed "break his balls" rompi coGLIOni (l. 72), but only when she is tired. Otherwise, she is a fantastic and calm woman. In this respect, he positions her weakness not as a personality flaw, but as a common human frailty. Overall, the confrontation is characterized by finishing each other's sentences on a verbal level and by gestures of loving connection on a bodily level.

5 Discussion

What all our data have in common is that long-married partners talk about themselves as a romantic couple. The flow of narration is interrupted every time one of the two partners chooses a formulation that might be face-threatening to the other. Sometimes, the delicate positioning of the other person pertains to the recounted past, and other times, it relates to the lived present. In all instances, however, it concerns a perceived deficit in the couple's relationship: "He" did not propose marriage (properly), "he" did not visit her frequently enough during their initial infatuation, "he" called her a "pain in the neck" 2 days after the wedding.

In all 48 cases analyzed, the epistemic stance-taking changes as a result of the side-sequence negotiation and the subsequent narrative builds on the slightly actualised story version. The spouses use the negotiation sequence to not only refresh their memory, but also humorously arrive at a story display in which both are in a good position. Moreover, they use this side sequence to publicly demonstrate that they can overcome difficulties in communication, even enjoying it, provoking each other a little and humorously dismissing the provocation.

The individual so positioned does not accept this characterization within the context of the "we-story". Typically, the discomfort is initially expressed bodily and subsequently articulated verbally. It manifests through more pronounced movements in turning toward and away from each other, in maintaining or breaking eye contact, and, in all cases, it is gestures that ultimately lead to reassurance and thus to the stabilization of the displayed relationship in the here and now of the joint story-telling in front of the camera.

It has been shown that the more positive, joint memory-making experiences a married couple has, the higher the levels of contentment within the marriage (Alea et al., 2015; Alea and Vick, 2010; Gildersleeve et al., 2017). Why do we-stories have such an impact on couples' happiness? Four main communicative purposes to we-stories were suggested by Singer and Skerrett (2014):

- We-stories help to name and structure the routines and values of the couple, thus putting the couple's identity into words.
- We-stories are a way of articulating meaning and purpose for the couple.
- We-stories act as a reminder of the love and commitment between the couple during conflict, thereby allowing for negotiation and growth.
- We-stories are a way of gathering and summarizing the wisdom and experiences of the couple in such a way that it can be transmitted to others.

In our data, we find three possible outcomes of the negotiation: (a) the partner may retract the counter-stance and agree on his wife's initial stance (cf. excerpt "Four times"), (b) the couple may achieve a new shared stance that integrates parts of both her initial stance and his counter-stance (this is what happens in the "Marriage proposal") or (c) the partner might slightly change the topic under discussion (as in the excerpt "She breaks my balls"). But no matter how the negotiation ends, the collaborative telling is characterized by an affective display of commitment (Tomasello, 2021) and—despite some initial trouble due to divergent stances—finally becomes a moment of embodied pleasure again (cf. Skerrett, 2013, 2016; Skerrett and Fergus, 2015).

Following, Barsalou et al. (2003) and Koch (2013) we understand the concept of "embodiment" as a constituent part of our being-in-the-world: "the body is there from the beginning and movement is what makes it perceptible in the first place" (Koch, 2013, p. 18). In this line of thought, and beyond the classical topic of inferring intention from motion, embodiment can refer to social stimuli (like a possibly face-threatening stance-taking) that cause the activation of bodily resources in counter-stances, but also to the crucial impact of the bodily movements on the sequential progressivity in interaction that are not only perceived by co-participants, but lead to renewed dialogic resonance.

Dialogic resonance (Du Bois, 2014; Zima, 2014) is created by speakers reusing parts of a previous utterance (as shown above in ex. 1 and 2) for activating the perception of similarity and thus connecting utterances that are not necessarily connected on the syntactic level. Building on Du Bois, Warner-Garcia (2013) and, in a similar vein, Chui (2014), transfer the concept of resonance to the analysis of gestures and identify two types of gestural resonance: *collaborative* and *problematising* resonance, both located in different sequential positions and concerning different communicative dimensions such as *co-semiosis* (excerpt "Marriage proposal"), *common ground* (excerpt "Four times"), and *coordination* (excerpt "She breaks my balls…"). In the stance negotiations under scrutiny here, dialogic resonance of a previous composite utterance may occur with an interesting amount of variation at different levels, resulting in varying degrees of similarity and contrast. One of the criteria for resonance is a kind of "active engagement" (Du Bois, 2014) with the previous utterance, as observed in our data. This engagement becomes visible in our analysis through the uptake of parts of an utterance by another speaker to perform their counter-stance (Zima, 2013) in a plethora of *intercorporal* engagements. Note that the narrative flow is interrupted at this point (Satti, 2023a); the partner who has been listening up to this point and who is affected by the face threat makes their own position heard. This opens a side sequence in which both partners ensure that the ascribed action is not concealed at the end, but that it is told in such a way that neither of them looks bad.

Overall, the spouses allow for side-sequences in which stances are negotiated; however, they do not allow for a real quarrel and thus a decisive break in the activity of collaboratively telling their story. Rather, one partner challenges an emergent stance and, subsequently, makes a counter-stance. Embodied resources in a wider sense and embodied resonance in a more specific sense come into play on two observational levels in our study on stance-taking. On the one hand as multimodal packages and on the other in the simultaneity of different statements on the verbal vs. the non-verbal level. In our data, multimodal packages are found as condensations of punchlines in story-telling: the "obvious-gesture" resonance in example 1 for the (un)clear marriage proposal, the resonant "list gestures" in example 2 for his too rare visits, the mutual touch for the challenge (and, ultimately, alteration) of the wife's stance-taking who presented herself as a pain in the neck in example 3. We conclude that the micro-sequential analysis made it possible to discover the highly complex interplay of multimodal resources like verbal and gestural resonance, mutual touch, synchronized position shifts, eye contact, choral vocalizations and, maybe most importantly, joint laughter. These multimodal resources help to "build new action by reusing with transformation [verbal and nonverbal] materials" inherited from prior speakers (Goodwin, 2018, p. 20), and thereby facilitate the romantic partners' co-operative achievement of shared epistemic and/or affective stance-taking in collaborative story-telling.

In a nutshell, investigating an Italian corpus of romantic couples' collaborative story-tellings of how they met, fell in love, proposed marriage, and quarrel as spouses, we found different sequential formats of multimodal packages and gestural resonances that contribute to negotiating epistemic, affective, and sometimes deontic stances. The overall goal seems to be to make a suddenly emergent rivalry friendly again, in order to make one's own individual voice heard, but still accomplish "doing being couple", i.e., performing the collaborative task as a loving "with" (Mondada, 2024) by "re-writing" the romantic couples' we-stories on the fly.

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 study was conducted in accordance with the local legislation and institutional requirements. The participants provided written informed consent to participate in this study. Written informed consent was obtained from the individual(s)

for the publication of any identifiable images or data included in this article.

Author contributions

SP: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing, Formal analysis, Resources, Supervision, Data curation. CP: Writing – review & editing, Methodology, Writing – original draft, Investigation, Data curation, Conceptualization, Formal analysis.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This article presents results from the interdisciplinary projects "Emergent Remembering. Fragmented Syntax and Textual Production in Contemporary Literature and Oral History" and "Saying the Unsayable", funded by the Schweizer Nationalfonds and the Deutsche Forschungsgemeinschaft (2018–2027, Project Number: 391351163, spokespersons: Thomas Klinkert and Stefan Pfänder), and the "Embodiment Research Network-EmbodiNet" (2024–2027, Project Number BA 7847/6-1, spokespersons: Prisca Bauer and Thomas Fuchs).

Acknowledgments

We thank two reviewers, and the editors for providing us with valuable feedback on an earlier version of this article. We are also grateful to Ignacio Satti, Daniel Alcón, and Andrea Knöbel for their conceptual input, Federica D'Antoni for recording and transcribing the data, Jonas Lüttke, Sarah Möller, and Anne Winterer for their efforts in refining the text and transcriptions quality and, not least, Philipp Nierle for the visualizations (screen shot drawings). All the remaining errors are, of course, our own.

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

Alea, N., Singer, J. A., and Labunko, B. (2015). ""We-ness" in relationship-defining memories and marital satisfaction," in *Couple Resilience Across the Lifespan—Emerging Perspectives*, eds K. Skerrett and K. Fergus (New York, NY: Springer), 123–137.

Alea, N., and Vick, S. C. (2010). The first sight of love: relationship-defining memories and marital satisfaction across adulthood. *Memory* 18, 730–742. doi: 10.1080/09658211.2010.506443

Andries, F., Meissl, K., de Vries, C., Feyaerts, K., Oben, B., Sambre, P., et al. (2023). *Multimodal Stance-Taking—A Systematic Literature Review*.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends Cogn. Sci.* 7, 84–91. doi: 10.1016/S1364-6613(02)00029-3

Biber, D., and Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text-interdiscipl. J. Study Discour.* 9, 93–124. doi: 10.1515/text.1.1989.9.1.93

Bressem, J., and Müller, C. (2014) "The family of Away Gestures: Negation, Refusal, and Negative Assessment," in *Body—Language—Communication. An International Handbook on Multimodality in Human Interaction (HSK 38.2)*, eds. C. M?ller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin/M?nchen/Boston: De Gruyter Mouton), 1592–1604.

Bröker, S., and Zima, E. (2022). Disaffiliierende Bewertungen und Haltungsbekundungen in Erzählaktivitäten—eine multimodale Analyse. *Ling. Online* 118, 29–55. doi: 10.13092/lo.118.9087

Brône, G., and Zima, E. (2014). Towards a dialogic construction grammar. *A corpus-based approach to ad hoc routines and resonance activation. Cogn. Linguist.* 25, 457–495. doi: 10.1515/cog-2014-0027

Chui, K. (2014). Mimicked gestures and the joint construction of meaning in conversation. *J. Pragm.* 70, 68–85. doi: 10.1016/j.pragma.2014.06.005

Clark, H. H. (1996/2012). *Using Language*. Cambridge. Cambridge University Press.

Clark, H. H., and Brennan, S. E. (1991). "Grounding in communication," in *Perspectives on Socially Shared Cognition*, eds. L. B. Resnick, J. M. Levine, and S. D. Teasley (Washington, DC: American Psychological Association), 127–149.

Couper-Kuhlen, E., and Selting, M. (2018). *Interactional linguistics: Studying Language in Social Interaction. Online Chapter C: Stance and Footing*. Cambridge: Cambridge University Press.

Deppermann, A., and Gubina, A. (2021). When the Body Belies the Words: Embodied Agency With darf/kann ich? ("May/Can I?") in German. *Front. Commun.* 6:661800. doi: 10.3389/fcomm.2021.661800

Deppermann, A., and Schmidt, A. (2021). Micro-sequential coordination in early responses. *Discour. Proc.* 58, 372–396. doi: 10.1080/0163853X.2020.1842630

Deppermann, A., and Streeck, J. (2018). *Time in Embodied Interaction*.Amsterdam; Philadelphia, PA: John Benjamins.

Du Bois, J. W. (2007). The stance triangle. *Stancet. Disc. Subject. Eval. Inter*. 164, 139–182. doi: 10.1075/pbns.164.07du

Du Bois, J. W. (2014). Towards a dialogic syntax. *Cogn. Linguist*. 25, 359–410. doi: 10.1515/cog-2014-0024

Enfield, N. J. (2013). A 'Composite Utterances' approach to meaning. *Anat. Mean Speech Mean. Comp. Utter. (Handbooks of Linguistics and Communication Science HSK*). 38, 689–707. doi: 10.1515/9783110261318.689

Freeman, V. (2019). Prosodic features of stances in conversation. *Lab. Phonol.* 10:163. doi: 10.5334/labphon.163

Gildersleeve, S., Singer, J. A., Skerrett, K., and Wein, S. (2017). Coding "We-ness" in couple's relationship stories: a method for assessing mutuality in couple therapy. *Psychother. Res.* 27, 313–325. doi: 10.1080/10503307.2016.1262566

Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse Soc.* 18, 53–73. doi: 10.1177/0957926507069457

Goodwin, C. (2018). *Co-operative Action*. Cambridge. Cambridge University Press.

Haddington, P. (2006). The organization of gaze and assessments as resources for stance taking. *Text Talk* 26, 281–328. doi: 10.1515/TEXT.2006.012

Hofstetter, E., and Keevallik, L. (2020). Embodied interaction. *Handb. Pragm. 23rd Annu. Installm*. 23, 111–138. doi: 10.1075/hop.23.emb2

Huber, J. R. (2015). Positive couple therapy: using we-stories to enhance resilience. *J. Marital Fam. Ther*. 41:251. doi: 10.1111/jmft.12091

Hügel, M. (2012). *Wenn Familien sich und andere(n) Geschichten erzählen. Die Narrative Konstitution von Familie durch gemeinsame Positionierungen in der Interaktion*. Freiburg: Albert-Ludwigs-Universität.

Kärkkäinen, E. (2006). Stance taking in conversation: from subjectivity to intersubjectivity. *Text Talk* 26, 699–731. doi: 10.1515/TEXT.2006.029

Kärkkäinen, E. (2007). The role of 'I guess' in conversational stancetaking. *Stancetak. Discourse Subject. Eval. Interact.* 183–219. doi: 10.1075/pbns.164.08kar

Kaukomaa, T., Peräkylä, A., and Ruusuvuori, J. (2014). Foreshadowing a problem: turn-opening frowns in conversation. *J. Pragm.* 71, 132–147. doi: 10.1016/j.pragma.2014.08.002

Koch, S. C. (2013). *Embodiment. Der Einfluss von Eigenbewegung auf Affekt, Einstellung und Kognition. Empirische Grundlagen und klinische Anwendungen*. Berlin: Logos.

Ladewig, S. (2014). "Recurrent gestures," in *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction (HSK 38.2)*, eds. by C. M?ller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressem (Berlin/M?nchen/Boston: De Gruyter Mouton), 1558–1574.

Ladewig, S. (2024). "Recurrent gestures: cultural, individual, and linguistic dimensions of meaning-making," in *The Cambridge Handbook of Gesture Studies. Cambridge Handbooks in Language and Linguistics*, ed. A. Cienki (Cambridge University Press), 32–55. doi: 10.1017/9781108638869.003

Marrese, O. H., Raymond, C., Wesley, B., Fox, B. A., Ford, C. E., and Pielke, M. (2021). The grammar of obviousness: the palm-up gesture in argument sequences. *Front. Commun.* 6:663067. doi: 10.3389/fcomm.2021.663067

Meyer, C. (2014). Menschen mit Demenz als Interaktionspartner. *Eine Auswertung empirischer Studien vor dem Hintergrund eines dimensionalisierten Interaktionsbegriffs. Zeitschrift für Soziologie* 43, 95–112. doi: 10.1515/zfsoz-2014-0203

Mondada, L. (2015). "Multimodal completions, " in *Temporality in Interaction*, eds A. Deppermann and S. Günthner (Amsterdam; Philadelphia, PA: Benjamins), 267–308.

Mondada, L. (2018). Multiple temporalities of language and body in interaction: challenges for transcribing multimodality. *Res. Lang. Soc. Interact.* 51, 85–106. doi: 10.1080/08351813.2018.1413878

Mondada, L. (2024). "Mobile body arrangements in public space: Revisiting "withs" as local accomplishments," in *New Perspectives on Goffman in Language and Interaction*, eds. L. Mondada, and A. Peräkylä (Routledge), 241–276.

Pekarek Doehler, S. (2019). At the interface of grammar and the body: Chais Pas ("dunno") as a resource for dealing with lack of recipient response. *Res. Lang. Soc. Interaction* 52, 365–387. doi: 10.1080/08351813.2019.1657276

Pfänder, S. (2023). Dimensionen der Leiblichkeit in der Interaktion. *Jahrbuch der Heidelberger Akademie der Wissenschaften*. 2022, 42–47.

Pfänder, S., and Couper-Kuhlen, E. (2019). Turn-sharing revisited: an exploration of simultaneous speech in interactions between couples. *J. Pragm.* 149, 13–30. doi: 10.1016/j.pragma.2019.05.010

Ruusuvuori, J., and Peräkylä, A. (2009). Facial and verbal expressions in assessing stories and topics. *Res. Lang. Soc. Interact.* 42, 377–394. doi: 10.1080/08351810903296499

Satti, I. (2023a). *Progresividad en la narración colaborativa*. Berlin: Un análisis multimodal. Peter Lang.

Satti, I. (2023b). Requests for verification across varieties of Spanish: a comparative approach to gaze behaviour. *Contrast. Pragm.* 1, 1–33. doi: 10.1163/26660393-bja10092

Schmid, H.-J. (2020). "Co-semiosis and other interpersonal activities," in *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment* (Oxford: Oxford Academic), 29–42.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., et al. (2009). *Gesprächsanalytisches transkriptionssystem 2 (GAT 2)*. Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion.

Singer, J. A., and Skerrett, K. (2014). *Positive Couple Therapy: Using We-Stories to Enhance Resilience*. New York, NY: Routledge.

Skerrett, K. (2013). "Resilient relationships: cultivating the healing potential of couple stories," in *Creating Connection: A Relational-Cultural Approach with Couples*, eds J. Jordan and J. Carlson (New York, NY: Routledge), 45–60.

Skerrett, K. (2016). We-ness and the cultivation of wisdom in couple therapy. *Fam. Proc.* 55, 48–61. doi: 10.1111/famp.12162

Skerrett, K., and Fergus, K. (2015). *Couple Resilience: Emerging Perspectives*. New York, NY: Springer.

Stevanovic, M. (2021). Three multimodal action packages in responses to proposals during joint decision-making: the embodied delivery of positive assessments including the finnish particle ihan "Quite." *Front. Commun.* 6:660821. doi: 10.3389/fcomm.2021.660821

Strong, T., Rogers-de Jong, M., and Merritt, S. (2014). Co-authoring "we-ness" and stories of intimacy. *Contemp. Fam. Ther.* 36, 398–408. doi: 10.1007/s10591-014-9304-8

Stukenbrock, A. (2021). Multimodal gestalts and their change over time: is routinization also grammaticalization? *Front. Commun.* 6:662240. doi: 10.3389/fcomm.2021.662240

Tanaka, S. (2016). Intercorporeality as a theory of social cognition. *Theory Psychol.* 25, 455–472. doi: 10.1177/0959354315583035

Tomasello, M. (2021). *Becoming Human. A Theory of Ontogeny*. Harvard: Harvard University Press.

Trujillo, J. P., and Holler, J. (2021). The kinematics of social action: visual signals provide cues for what interlocutors do in conversation. *Brain Sci.* 11:996. doi: 10.3390/brainsci11080996

Warner-Garcia, S. (2013). Gestural resonance: the negotiation of differential form and function in embodied action. *Crossroads Lang. Inter. Culture* 9, 55–78.

Yang, M., and Wang, M. (2025). Recurrent gestures and embodied stance-taking in courtroom opening statements. *Text Talk* 45, 273–299. doi: 10.1515/text-2023-0042

Zima, E. (2013). *Kognition in der Interaktion. Eine kognitiv-linguistische und gesprächsanalytische Studie dialogischer Resonanz in österreichischen Parlamentsdebatten*. Heidelberg: Winter.

Zima, E. (2014). Gibt es multimodale Konstruktionen? Eine Studie zu [V (motion) in circles] und [all the way from X PREP Y]. *Gesprächsforschung–Online-Zeitschrift zur verbalen Interaktion.* 15, 1–48.

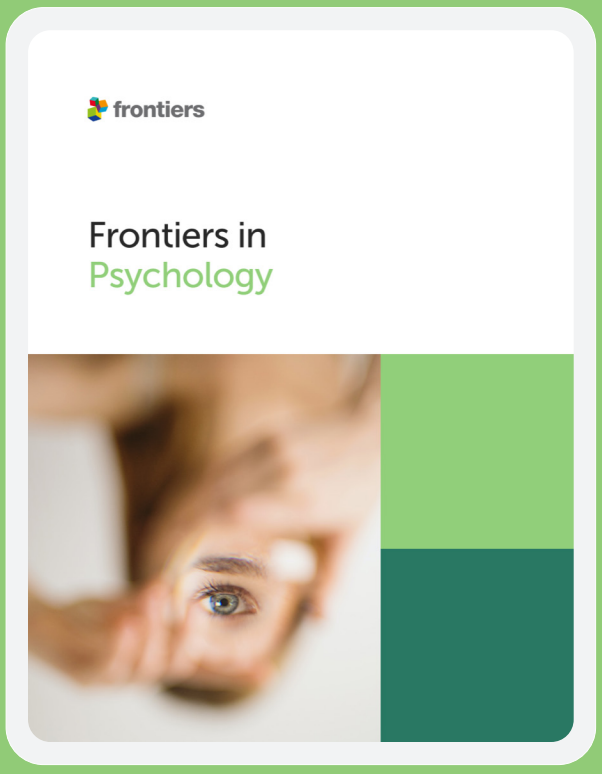
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

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