



OPEN ACCESS

EDITED BY

Marco De Angelis,
University of Bologna, Italy

REVIEWED BY

Patrik Pluchino,
University of Padua, Italy
Luca Gualtieri,
Free University of Bozen-Bolzano, Italy

*CORRESPONDENCE

Franziska Legler
✉ franziska.legler@mb.tu-chemnitz.de

RECEIVED 21 July 2025

ACCEPTED 08 October 2025

PUBLISHED 05 November 2025

CITATION

Legler F and Bullinger AC (2025) How to achieve human-centered automation: the importance of trust for safety-critical behavior and intention to use in human-robot collaboration.

Front. Organ. Psychol. 3:1669782.
doi: 10.3389/forgp.2025.1669782

COPYRIGHT

© 2025 Legler and Bullinger. 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.

How to achieve human-centered automation: the importance of trust for safety-critical behavior and intention to use in human-robot collaboration

Franziska Legler* and Angelika C. Bullinger

Chair of Ergonomics and Innovation, Department of Mechanical Engineering, Chemnitz University of Technology, Chemnitz, Germany

Introduction: Recent technological advances in human-robot collaboration (HRC) allow for increased efficiency and flexibility of production in Industry 5.0 while providing a safe workspace. Despite objective safety, research has shown subjective trust in robots to shape the interaction of humans and robots. While antecedents of trust have been broadly examined, empirical studies in HRC investigating the relationship between trust and industry-relevant outcomes are scarce and the importance of trust regarding its precise effects remains unclear. To advance human-centered automation, this paper investigates the affective, cognitive, and behavioral consequences of trust in robots, and explores whether trust mediates the relationship between industry-relevant characteristics and human-centered HRC outcomes.

Methods: In a pseudo real-world test environment, 48 participants performed a manufacturing task in collaboration with a heavy-load robot. Trust, affective experience over time, intention to use, and safety-critical behavior were examined. A $2 \times 2 \times 2$ mixed design varied the availability of feedback, time pressure, and system failures, each expected to affect the level of trust.

Results: In the control group, trust remained consistently high across all conditions. System failures and feedback significantly reduced trust, whereas time pressure had no effect. System failures further increased negative affective experience, while feedback reduced safety-critical behavior. Trust was unrelated to affective experience but positively related to safety-critical behavior and intention to use. The relationship between feedback and safety-critical behavior, as well as intention to use, was significantly mediated by trust.

Discussion: Highly relevant for implementation, the control group showed a tendency toward overtrust during collaboration, evidenced by disregarding system failures. The results indicate that implementing a feedback system alongside the simulation of safe system failures has the potential to adjust trust toward a more appropriate level, thereby reducing safety-critical behavior. Based on these findings, the paper posits several implications for the design of HRC and gives directions for further research.

KEYWORDS

human-robot collaboration, trust in automation, affective experience, safety-critical behavior, intention to use, human factors, manufacturing

1 Introduction

Due to demographic changes and the declining prestige and attractiveness of workplaces involving physical labor among younger generations, meeting personnel demands in manufacturing companies is becoming an increasingly critical challenge for success. Furthermore, manufacturing companies experience an increasing demand for customized products. Economical production of small batch sizes and variant flexibility have become key to success (Buxbaum and Sen, 2021). Conventional automation reaches its limits here. However, technological advancements within Industry 4.0 like adaptive sensor systems (e.g., Bdiwi et al., 2017a) enable humans to interact with robots in a shared workspace without safety fences and to work simultaneously on the same task or component (ISO International Organization for Standardization, 2011). This so-called human-robot collaboration (HRC) is directly associated with Industry 5.0 (Alves et al., 2023) and is expected to unlock significant potential in the productive industry: (i) combining the repetitive skills of robots to improve quality with the ability of humans to solve ill-defined problems (ISO International Organization for Standardization, 2016), (ii) improving flexibility of production facilities (Oubari et al., 2018), (iii) improvement of ergonomic work conditions, both physical and mental, as robots execute monotonous or heavy-load tasks (Buxbaum and Sen, 2021; Pietrantoni et al., 2024), and (iv) simple programming and operation, enabling operators without programming expertise to train robots (Buxbaum and Sen, 2021). However, shared-space interaction with robots is still rarely implemented on the shop floor (Kopp et al., 2021; Giallanza et al., 2024). In addition to the technical complexity, this is partly due to the expectation that HRC holds a potential risk for humans (Buxbaum and Sen, 2021) as malfunctions of robots are even possible in standardized collaborations in industrial manufacturing (Schäfer et al., 2024). To both overcome these concerns and realize the potential of HRC, Industry 5.0 focuses on human-centered automation, sustainability, and resilient production, instead of the narrower “techno-economic vision” (Alves et al., 2023), which aims to maximize production speed and economic profit. Thus, the consideration of human factors is now central in HRC (Alves et al., 2023; Giallanza et al., 2024; Paliga, 2023), as is the understanding of the experience of humans during interactions with intelligent automation, such as robots (Alves et al., 2023).

From a human factors perspective, trust in robots is considered a key factor, as it determines the way humans interact with the robot (Coronado et al., 2022; Hancock et al., 2021; Schäfer et al., 2024). One of the most frequently cited contributions in the field of trust in HRC is the review paper of trust in automation by Lee and See (2004) – written at a time when HRC was still a young domain. The authors emphasize that, when interacting with automation, trust must be calibrated appropriately, as under- and overtrust are both associated with negative outcomes such as the misuse, disuse and abuse of automated systems. To this day, research emphasizing the importance of trust in HRC still largely refers to this foundational work (e.g., Hancock et al., 2011; Hopko et al., 2023)—although Lee and See (2004) primary focused on information and decision support automation rather than robots. Two recent reviews of

shared-space HRC (Hopko et al., 2022; Coronado et al., 2022) concluded that the characteristics of robots as antecedents of human factors such as trust are widely studied. They have also shown that human factors are only used as dependent variables whereas their interrelations and potential effects on other covariate human factors as well as their effects on further industry-relevant outcomes in a shared-space HRC have so far been largely neglected. Given that research indicates trust to play a central role in HRC (e.g., Hancock et al., 2021; Schäfer et al., 2024), it seems pertinent that trust should be examined more closely as a predictor of industry-relevant outcomes. Therefore, in addition to studying the antecedents of trust in HRC, research in the field needs to further investigate the actual consequences and outcomes of trust relevant to the shop floors of Industry 5.0. Industry-relevant outcomes can be divided into two categories: *performance-centered* and *human-centered*. Industry 4.0 focused primarily on performance-centered outcomes (e.g., productivity, production costs, resource efficiency, product customization) of technological advancements. Beyond that, Industry 5.0 emphasizes the importance of human-centered outcomes like wellbeing and human safety, which are key to achieving the goals of Industry 4.0. Investigating human-centered outcomes empirically will help to nuance the discussion in the field and contribute to the realization of HRC, overcoming the tendency to argue for the importance of trust mainly in light of earlier results gained from the more general field of trust in automation. When working toward the goal of human-centered automation, it is decisive to look beyond performance criteria and consider the affective, cognitive and behavioral consequences for the individual human in a shared-space HRC.

This paper contributes to the state of the art with an empirical study investigating the relationships with and potential mediating effects of trust in robots during HRC on affective, cognitive, and behavioral human-centered outcomes. The following chapter provides an overview of the state of the art regarding the concept of trust in robots in HRC, specific human-centered outcomes of trust in HRC for Industry 5.0, and antecedents of trust that can be used as strategies to adjust trust levels.

2 State of the art

Despite an extensive body of research in the field, there is still no consensus how trust is defined (Hopko et al., 2022). Proposing an inclusive, human-focused definition, Schäfer et al. (2024) “define human trust in robots as an asymmetric way of relating that is characterized by dependence, risk, vulnerability, positive expectations, and free choice (p. 9).” Transferring this definition to industrial HRC, it follows that humans (i) depend on the robot to fulfill their work task (dependence), (ii) perceive a risk as they cannot predict future actions of the robot (risk), (iii) are vulnerable due to potential physical harm or a lack of knowledge regarding the robot’s capabilities (vulnerability), (iv) expect the robot to support the shared task (positive expectation), and (v) are free to engage in trust toward the robot (free choice) even if they are obliged to use it in the workplace. In the same manner that it is possible to interact with another person without establishing trust, it is also possible to interact with a robot without establishing trust.

Trust, which is based on positive expectations, leads to different behavior than mistrust in both human-human and human-robot constellations (Schäfer et al., 2024). For example, as with human-human interactions, humans interacting with robots can choose to observe their interaction partner's every physical or mental move in the case of mistrust, or choose not to monitor them based on positive expectations in the case of trust. Therefore, in HRC, humans have a certain degree of freedom on how they interact with the robot.

In a specific situation, humans choose to trust or choose not to trust (Schäfer et al., 2024). This situational aspect of the definition by Schäfer et al. lead to the conclusion that trust levels differ depending on situational circumstances and context. Firstly, this mirrors the argument that an “appropriate” or “calibrated” level of trust is necessary for effective HRC (Lee and See, 2004; Hancock et al., 2011, 2021; Chang and Hasanzadeh, 2024). Secondly, trust levels can be adjusted depending on situational circumstances and context (Hancock et al., 2011).

Going beyond antecedents of trust, Schäfer et al. (2024) also include affective (dependence and vulnerability), cognitive (positive expectations) and behavioral (free choice) components. It can be concluded that trust should have consequences on all three components, respectively. This aligns with the human-centered focus of Industry 5.0, where individual wellbeing is prioritized (Alves et al., 2023; Giallanza et al., 2024).

On an *affective level*, subjective wellbeing is highly relevant in human-centered automation. Negative emotions and stress experienced at the workplace impact employees' wellbeing (Khalid and Syed, 2024). However, affective experience remains understudied in HRC (Coronado et al., 2022). When affective experience is addressed, anxiety is a commonly studied affective state in shared-space HRC (Hopko et al., 2022). In a comprehensive review, Mauss and Robinson (2009) conclude that there is no “gold standard” for measuring emotions, as existing methods fail to capture both subjective experiences and physiological or behavioral reactions. The authors advocate for a dimensional approach to affective assessment, emphasizing valence and arousal as essential components. They also stress the importance of measuring emotions in conjunction with significant events to ensure accurate results. According to the circumplex model of affect (Russell, 1980), negative affective experiences are characterized by negative valence and increased arousal. To the best of our knowledge, there are no theoretical models that explicitly consider the relationship of trust and affective experience. Therefore, it is still unclear whether trust is an antecedent, covariate or consequence of affective states, and studies that track affective states over time and in relation to specific events are scarce in the context of HRC. As both trust and affective experience are examined by changes in physiological measures (e.g., Arai et al., 2010; Hopko et al., 2022), it can be assumed that both concepts are related. Based on trust literature, it has been found that (i) well-studied factors influencing trust (e.g., reliability) also affect physiological responses (Hopko et al., 2023), (ii) (3) emotions were suggested to be useful markers for under- and overtrust (Schoeller et al., 2021), and (iii) humans react with negative affective states following trust violations by robots (Alarcon et al., 2024). Applying the definition of Schäfer et al. (2024), if humans show low trust in a robot, it can be assumed

that feelings of “vulnerability” arise, which might be accompanied by negative affective experiences such as anxiety—even if the robot performs perfectly. From these empirical and theoretical findings, we deduce that trust is an antecedent for affective experiences. Therefore, we exploratorily hypothesize:

Hypothesis 1: Lower trust is related to reduced valence and increased arousal (representing negative affective experience) during HRC.

On a *cognitive level*, for HRC to be implemented sustainably and resiliently in industrial workplaces, human workers should have a willingness to use the technology and have positive expectations of it. This is expected to be related to their belief that a certain type of automation, e.g., a robot in HRC, is beneficial to their goals (cf. “positive expectations,” Schäfer et al., 2024). However, trust is distinct from both acceptance and actual utilization of technology. While trust can enhance both, they have been found to be strongly influenced by contextual moderating factors (Kopp, 2024). A recent meta-analysis suggests that trust is actually an antecedent of intention to use a robot (Razin and Feigh, 2024). Additionally, trust has been shown to mediate the relationship between usability and willingness to use a robot (Babamiri et al., 2022). Therefore, intention to use is a relevant criterion, influenced by positive expectations as part of trust. However, most studies on trust in automation have focused on the relationship between automation reliability and operator usage, often neglecting to assess trust as a mediating variable (Lewis et al., 2018). We intend to further knowledge in this area and hypothesize:

Hypothesis 2: Trust is positively related to intention to use.

On a *behavioral level*, safety is paramount for human-centered automation (cf. “risk” and “vulnerability,” Schäfer et al., 2024). While shared-space HRC increases the risk of being physically harmed by the robot (Onnasch and Hildebrandt, 2022), it has been found that people are subject to a “positivity bias” (Dzindolet et al., 2003): they expect automated systems to work well and reliably. A recent study of technical experts in HRC even found concern about potential over-reliance on technology (Pietrantonio et al., 2024). The resulting overtrust reduces situation awareness and is related to complacency (Parasuraman and Manzey, 2010). On an objective and observable behavioral level, higher trust has been found to result in reduced monitoring behavior, with an increase in trust during interaction leading to even less monitoring (Hergeth et al., 2016). Therefore, it was shown that humans objectively react differently during interactions depending on their level of trust (cf. “free choice,” Schäfer et al., 2024). In example, workers might stop monitoring a cobot's actions in case of too high trust (Hopko et al., 2022). As malfunctions are rare and trust increases over time during failure-free interaction, consequently reduced monitoring may lead to safety-critical situations. Workers may, due to reduced monitoring, be unaware of system states and unable to react to sudden off-nominal events (Hopko et al., 2022) like system failures. This concept of overtrust in robots has been defined as a situational state in which humans misjudge the risks associated with their own actions due to an underestimation of the likelihood of a robot performing its functions ineffectively or unsafely (Wagner et al., 2018; Aroyo et al., 2021). Therefore, it is important to

better understand objective safety-critical behavior of humans as a behavioral consequence of trust. Concluding, we hypothesize:

Hypothesis 3: Trust is positively related to safety-critical behavior during shared-space industrial HRC.

As mentioned above, the body of knowledge on antecedents of trust is extensive (Hopko et al., 2022). Previous results point to differences in the antecedents of trust depending on the domain. For example, although anthropomorphism has generally been found to have a positive impact on trust in robots, this holds particularly true for social robots and not necessarily for industrial robots (Onnasch and Hildebrandt, 2022). Research in the domain of industrial HRC has thus to consider industry-relevant characteristics as antecedents of trust, as proposed e.g., by Chang and Hasanzadeh (2024), who underline the need to investigate distinctive features relevant to the respective field of robots. The results of trust-influencing factors can be further used to create variance in human trust levels by purposefully adjusting trust levels and subsequently explore relationships with expected consequences of trust. In industrial application, two robot characteristics—*system failures* and *feedback* via communication—and the environmental characteristic *time pressure* have been argued to be important trust antecedents (Chang and Hasanzadeh, 2024). The following section briefly describes the effects of these antecedents on trust. Based on the previously explained relationship between trust and human-centered outcomes in industry, the links between antecedents of trust and these outcomes are discussed.

Much of the literature on trust focuses on system reliability and the effects of system failures; the latter have been shown to have a negative impact on trust (e.g., Parasuraman and Manzey, 2010; Wickens et al., 2015; Legler et al., 2023). This can be explained by the definition of trust itself: As trust involves positive expectations, these are violated in the event of a system failure. Akalin et al. (2023) state that trust loss is especially caused by robot failures, which include design failures, expectation failures and system failures. In this paper, we follow Akalin et al. (2023) and define system failures as failures in both, hardware or software, of a robot system. Furthermore, system failures increase perceptions of vulnerability (Schäfer et al., 2024). Higher monitoring and situation awareness have been observed following failures of an automated system (Wickens et al., 2015). As trust is positively related to intention to use and safety-critical behavior, it can be assumed that system failures also reduce these. Further, as trust has been shown to be related to affective experience (e.g., Alarcon et al., 2024; Hopko et al., 2023), it can be assumed that unexpected system failures are related to negative affective experience, probably mediated by trust. Given that studies on affective experiences are rare, the affective experiences resulting from trust-related antecedents and consequences of trust will be assessed in an exploratory manner. On the account of system failure and intention to use, we hypothesize:

Hypothesis 4: System failures reduce trust during industrial HRC, and thereby intention to use.

Providing *visual feedback* on the state of an automated system, such as a robot, has been shown to increase situation awareness and the predictability of robots' actions which enhances trust, thereby strengthening intention to use and reducing safety-critical

behavior. Visual feedback systems have been implemented to increase user's situation awareness (Maurtua et al., 2017), displaying visual cues to signal danger and capture human attention (Goldstein, 2010). A meta-analysis showed that feedback directly affects trust in automation (Schäfer et al., 2016). Additionally, trust mediates the relationship between feedback and actual reliance on the system (cf. "dependence," Schäfer et al., 2024), as feedback renders the robot's actions more predictable (Dzindolet et al., 2003). This minimizes human vulnerability, while increasing acceptance and intention to use. Therefore, although feedback is expected to enhance trust, and trust is positively related to safety-critical behavior, it is expected that feedback will lower safety-critical behavior. To better understand these relations, we hypothesize:

Hypothesis 5: Feedback increases trust during industrial HRC.

Hypothesis 6: Trust mediates the relationship of feedback with intention to use and safety-critical behavior.

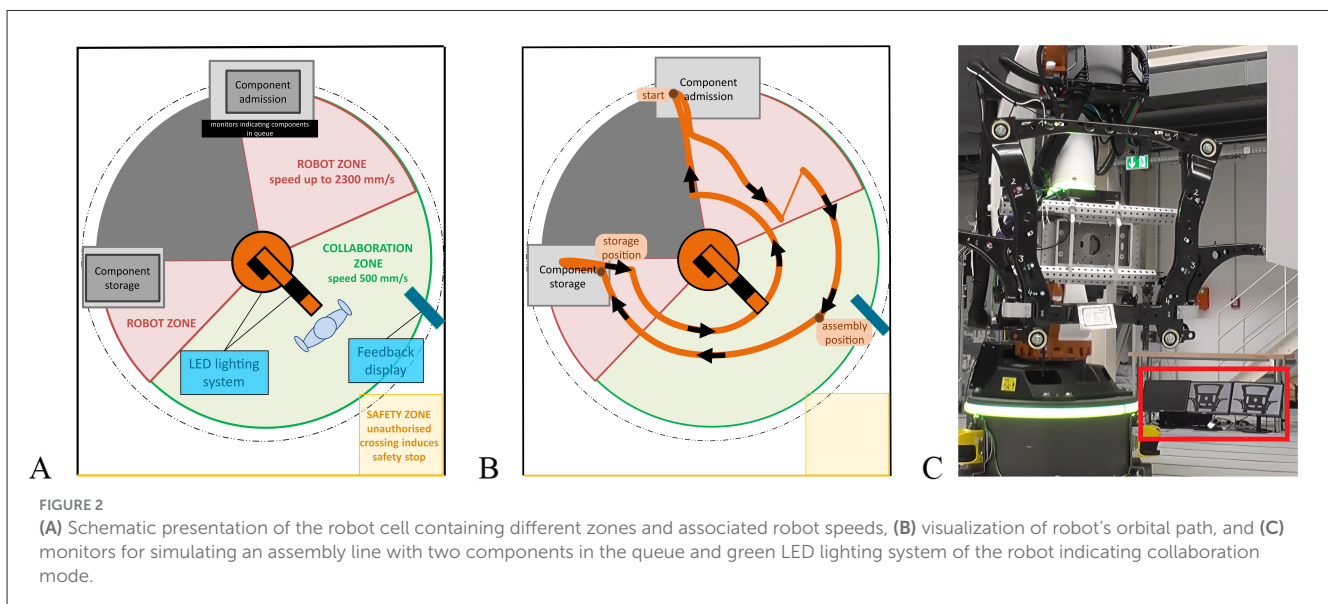
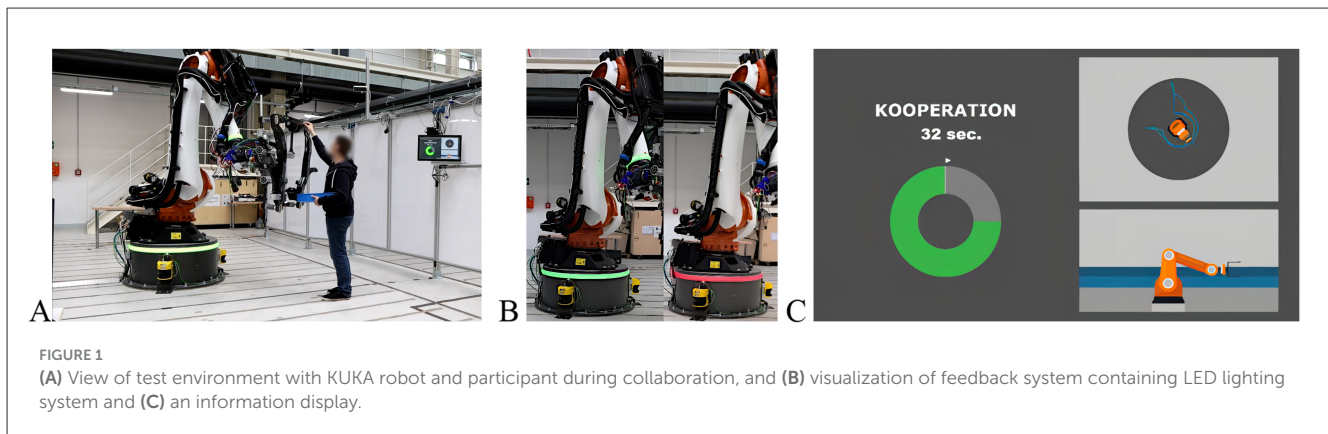
Hypothesis 7: Feedback shows a direct effect on reducing safety-critical behavior.

Pacing in industrial settings can cause *time pressure*, which has been found to increase mental workload during human-automation interactions. The framework of Chang and Hasanzadeh (2024) highlights time pressure as a key environmental factor affecting trust in robots in industrial settings. Surprisingly, effects of time pressure have remained understudied. Most research assumes that time pressure increases trust, as humans experience an increased subjective workload under time pressure (Wang et al., 2016) and compensate this complexity and uncertainty by increasing trust in automation (Luhmann, 1979; van der Waaij et al., 2021). As trust is positively related to intention to use and safety-critical behavior alike, it can be assumed that time pressure enhances both due to increased positive expectations and dependence on the robot. Therefore, we hypothesize:

Hypothesis 8: Time pressure increases trust during industrial HRC.

Hypothesis 9: Trust mediates the relationship of time pressure with intention to use and safety-critical behavior.

Summing up the current state of trust research, three elements are apparent that require closer consideration: (i) The concept of trust in HRC has characteristics at affective, cognitive and behavioral levels which leads to the assumption that trust also manifests in affective, cognitive, and behavioral consequences relevant for the design of efficient HRC in industrial settings. (ii) Investigating the concepts of affective experience, intention to use and safety-critical behavior is highly relevant to achieve human-centered automation as intended within Industry 5.0. (iii) Trust has the potential to mediate the relationship between characteristics typically apparent in industrial robots, such as system failures, feedback and time pressure, and important human-centered outcomes of industry, such as affective experience, intention to use and safety critical behavior. This puts trust at the center of attention to explain relationships between industrial characteristics of HRC workplaces and affective, cognitive, and behavioral consequences. Until now, human factors such as trust have rarely been studied as covariates of other human factors, and the potential mediating effects on outcomes relevant for a safe HRC have been largely neglected (Hopko et al., 2023). Therefore, our overall research



question is: In an industrial HRC setting, how do changes in trust, prompted by the trust-relevant antecedents system failures, feedback and time pressure, impact the human-centered outcomes affective experience, intention to use and safety critical behavior?

To address the research question, an experimental study was conducted, comprising a realistic industrial collaboration task with a heavy-load robot within a pseudo real-world test environment. Method and results of the study are subsequently presented, closing with practical implications for industrial HRC.

3 Materials and methods

3.1 Test environment

The study was conducted in a model factory. The robot cell in which participants operated was surrounded by various other robotic and automation systems. A KUKA Quantec prime KR 180 industrial robot, classified as a heavy-load robot, was used as a test bed with a feedback system implemented (see Figure 1A). The visual feedback system contained an LED lighting system at the flange and base of the robot (see Figure 1B) and an

information display (see Figure 1C). Three colors showed different robot states: white for automated robot actions, green for the collaboration mode, and orange for system failure states. The dynamic information display contained a countdown showing the remaining collaboration time in seconds, as well as top and side views of the robot to help anticipate actual and future movement and paths. There was also a pop-up window providing written information on failure states. The information display was mounted at an angle on the wall that ensured visibility from both the safety zone and the assembly position. Still, from the assembly position, the human operator had to turn sideways and away from the robot to see the entire display (see Figure 2A). Redundant information on the robot's state and failure modes was presented by the LED lighting system and the feedback display. Further details regarding the development and user evaluation of the feedback system can be found in Legler et al. (2022).

Table 1 describes the tasks performed by the human operator and the robot within a complete assembly cycle, as well as the information provided by the feedback system. Each cycle contains five phases involving the robot: component admission and delivery (phase 1), gesture control (phase 2), assembly (phase 3), component storage (phase 4), and return to component admission (phase 5).

TABLE 1 Robot and human tasks during one complete assembly cycle of the HRC task.

Robot task	Human task	Feedback system
Phase 1: Component admission and delivery (11 s)		
Admission of component (front axle carrier), transport toward and stop inside collaboration-zone	Waiting in safety zone, possibly sorting hook-and-pile tapes if necessary	Mode of operation (autonomous), current and future position of robot, countdown till collaboration
Phase 2: Gesture control (direct interaction, movement of robot, individual duration)		
Holding component on robot flange, support human's input by gesture control (speed 250 mm/s)	(1) Entering collaboration-zone (2) Gesture control: setting an ergonomic height of the component located at the robot flange by moving palm up and down	Mode of operation (collaborative), countdown till collaboration ends, warning and explanation in case of a system failure
Phase 3: Assembly (no movement of robot, individual duration)		
Holding component during assembly (no movement)	(3) Assembly of eight hook-and-pile tapes on the component (4) Entering safety-zone	Mode of operation (collaborative), countdown till collaboration ends
Phase 4: Component storage (10 s)		
Transport of component to storage position (departing from human), placement of the component	–	Mode of operation (autonomous), current and future position of robot
Phase 5: Return to component admission (14 s)		
Returning to component admission (approaching human)	–	Mode of operation (autonomous), current and future position of robot

Phases *gesture control* to enable an ergonomic height and *assembly* can be referred to as “collaboration,” as participants left the safety zone during these phases. Figure 2B illustrates the robot's orbital path. The robot: (i) picked up the front axle carrier at the back; (ii) approached toward the assembly position in the front; (iii) remained in the assembly position holding the front axle carrier for human gesture control and assembly; (iv) moved to the component storage position on the left, and (v) returned to home position in the back.

The study was conducted following the approval of a formal risk assessment. The robot cell comprised collaboration, robot and safety zones (see Figure 2A, cf. Bdiwi et al., 2017a for further details on safety concept in the robot cell). Participants remained inside the robot cell throughout the experiment, waiting in the safety zone until the collaboration mode began (otherwise triggering an emergency stop). The robot's speed was associated with particular zones; it moved at a maximum speed of 2,300 mm/s in the robot zone and slowed down to 500 mm/s when entering the collaboration zone (see Figure 2B) to ensure that the robot system was able to immediately stop in case of an emergency stop triggered by humans. The collaboration mode only started when the robot came to a standstill within the collaboration zone at the assembly position. The LiDAR sensors (see Figure 1A) were deactivated during collaboration mode. The robot was capable of supporting a collaborative assembly task modeled on a real automotive industry workplace. The task involved attaching eight hook-and-pile tapes to a front axle carrier. The robotic system was equipped with an optical gesture control interface for height adjustment which is a typical ergonomic feature in HRC (Pluchino et al., 2023). This interface is based on a combination of depth, color and thermal imaging (for technical details cf. Al Naser et al., 2022). In practice, the algorithm performed best when participants extended their palm toward the robot. Movements of the palm in the vertical direction relative to the initial detection position were then

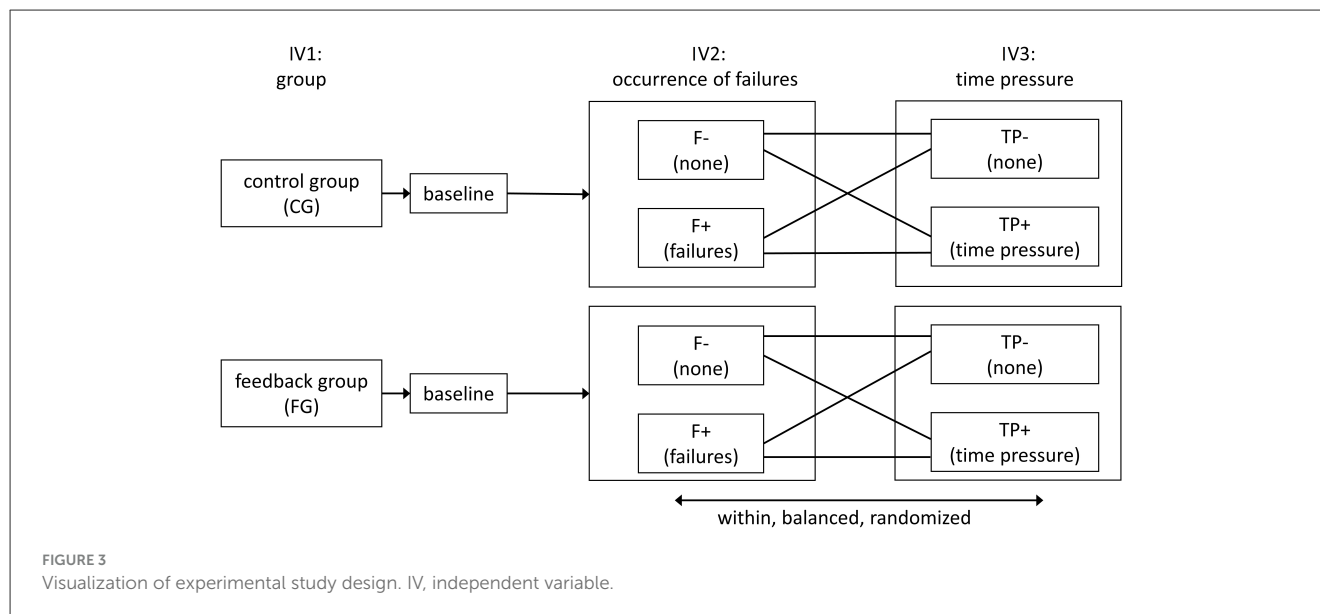
translated into corresponding upward or downward motions of the robot arm. During collaboration mode, the robot moved at a speed of 250 mm/s. Given the robot movement with minimal physical distance to the human in the collaboration zone to support the human height setting task and holding the component for assembly represents a HRC level 3 according to Bdiwi et al. (2017b).

Although the automated gesture control was functional, the gesture control algorithm sometimes failed in prior study series as a result of incorrect human input or adverse environmental influences such as light reflections. Algorithm failures resulted in a lack of response from the system or jerky, uncontrolled movements of the robot arm due to repeated failed tracking of the palm. Therefore, to ensure standardized conditions, the test environment enabled the experimenters to manually control the robot arm during gesture control. Participants were unable to detect the manual gesture control (Wizard-of-Oz), allowing experimenters to simulate a failure of the gesture control system. As the automated gesture control was based on a very simple human gesture, moving the extended palm upwards and downwards was equally used as the input gesture for the Wizard-of-Oz gesture control.

To simulate an assembly line, the test environment also included three monitors positioned behind the robot and beneath the component admission table (see Figure 2C). The monitors were in the person's direct field of vision during the assembly task, ensuring that participants could evaluate their performance in the current assembly.

3.2 Experimental design

A mixed design was conducted with three factors: feedback, system failure and time pressure. “Feedback” was a between-subjects factor (CG = control vs. FB = feedback group); therefore, in the control group, the LED lighting system and information



display were kept off. A balanced design of two within-subjects factors, “system failures” (F− = none vs. F+ = occurrence of failures) and “time pressure” (TP− = none vs. TP+ = time pressure), was applied within both groups. The experimental design resulted in four conditions, which were performed in randomized order after a baseline condition to allow habituation to the assembly task and to the assessment of affective experience. The design of the study is visualized in [Figure 3](#).

System failures were simulated in gesture control mode while participants were setting the assembly height. As the robot delivered the front axle carrier at a height that prevented participants from reaching the highest hook-and-pile tapes, participants’ intention was to lower the robot arm. The experimenters were instructed to manually control the robot arm with noticeably jerky movements in the direction opposite to the participants’ palm movements for about 10 s, resulting in a position far too high to reach all hook-and-pile tapes. Afterwards, gesture control was supported correctly and participants were able to lower the robot arm. Time pressure was introduced by adding a component to the assembly line monitors (see [Figure 2C](#)) after an individual cycle time had elapsed, resulting in a component in queue visible for the participants. Therefore, the time taken for collaboration mode, including time for gesture control and assembly of the component, was measured during participants’ baseline assembly cycles, and the fastest collaboration time was used as an individual reference for each participant. The constant robot period (35 s) and the individual reference time ($M = 26.8$ s, $SD = 4.7$, $min = 20$, $max = 39$) were added together and used as the individual cycle time (resulting in an average of 61.8 s). To account for interindividual differences in assembly times and to ensure that time pressure was objectively operationalized and equally perceived across participants, the individual reference time was halved during time pressure conditions. Therefore, individual reference time during time pressure conditions showed an average of 13.4 s, resulting in an average individual cycle time of 48.4 s. From the participants’ perspective, the individual reference time

showed them how much time they had available to collaborate with the robot, and this time was displayed on the information display (see [Figure 1B](#)).

3.3 Measures

3.3.1 Demographics

Demographic information like sex, age, highest educational degree, as well as experience with industrial robots and production work, were captured in pre-survey. Additionally, Affinity for Technology Interaction ([Franke et al., 2019](#), ATI, 6-point Likert-scale; $\alpha = 0.85$) was assessed for sample specification.

3.3.2 Time progression measures of affective experience

According to [Werth and Förster \(2015\)](#), specific emotions are intensive, directed toward an object, and fleeting. Since collaborations with robots involve alternating phases of separate and cooperative activities, and robots approach and depart from humans, various affective states can occur. In accordance with [Mauss and Robinson \(2009\)](#), a dimensional self-report measure of affect over time was used, as it has been argued that this is sensitive to changes in valence and arousal, whereas physiological measurements such as heart rate or skin conductance have only been shown to be sensitive to arousal. The circumplex model of affect ([Russell, 1980](#)), for example, describes anxiety as an emotion characterized by a strong negative valence and high arousal. Therefore, it is possible to assess a tendency toward specific emotions based on valence and arousal values. The Feeling Scale (FS; [Hardy and Rejeski, 1989](#); scale range −5 to 5) and Felt Arousal Scale (FAS; [Svebak and Murgatroyd, 1985](#); scale range 1 to 12) were used to quantify self-reported affective experience. Combined single-item scales have also been used in other studies to measure affect continuously over time ([Murray et al., 2016](#)). Valence and

arousal values were assessed every 10 seconds and the original scale of FAS was doubled to 12 points due to limited variance in values in pre-tests (cf. Legler et al., 2023 for further details on pre-test). Due to pretest results, the assessment interval of 10 seconds was found to enable participants to concentrate on their primary task. Both scales were available to the participants at all times, visualized by a fabric sticker on their left or right arm.

3.3.3 Outcome measures (post-scenario)

Further, several measures for dependent variables were applied. **Trust in automation** was measured subjectively via a German translation (Pöhler et al., 2016) of Jian-Scale, following each of the five scenarios. Only subscale “trust” (6 items; 7-point Likert scale) was used because the two subscales were highly correlated and previous work had questioned the two-factor structure (Pöhler et al., 2016; Legler et al., 2020). Across the scenarios, mean reliability for trust was $\alpha = 0.82$. For *Intention to Use*, a German translation of the subscale “intention to use” (Brauer, 2017) was applied (scale range 1 to 5, mean reliability was $\alpha = 0.78$) to measure overall willingness to interact with the robot in the future (cf. UTAUT Model; Brauer, 2017), representing the pre-conditional stage of actual usage and an indicator of technology acceptance. *Safety-critical behavior* was defined as disregarding system failures. After completion of data collection, the time taken for gesture control across all scenarios without simulated system failures was empirically calculated and showed a 5th percentile of 0.5 s. This value was used as a lower limit of an expected time for gesture control. Therefore, in failure scenarios, a minimum gesture control time of 10.5 s was expected (10 s for the failure simulation plus an expected lower limit of 0.5 s for the actual height adjustment). Due to the manual control of failures and expected slight inaccuracies of experimenters in simulating the failure duration, a tolerance period of 2 s was set. Therefore, in failure scenarios, a duration of less than 8.5 seconds between the start of collaboration mode and the start of assembly was defined as disregard of failures. De facto, participants did not complete the height setting task and started to assemble while the robot was still moving, which was not permitted by the instructions given to participants. To control for the *manipulation of time pressure*, one item covering temporal demands of the NASA-TLX measuring subjective workload (Hart, 2006; scale 0–100) was applied. Additionally, the number of components in the queue after finishing all five assembly cycles of a scenario was recorded for manipulation check.

3.4 Sample

The participants were recruited from the participant database of our partner research institution, which includes students from technical degree programmes and individuals working in technical professions, e.g., employees from companies located in the vicinity. Participants in the study had to be at least 18 years and free from any motor impairments. Given the technical focus of the study, an above-average affinity for technology was expected. In the experiment, 48 persons participated and were randomly assigned to two groups of equal size: control group and feedback group. Both

groups were gender balanced. Participants were aged between 18 and 58, the mean age was 26.2 years ($SD = 7.54$) in the control and 25.6 years ($SD = 7.84$) in the feedback group. Bayes factors (cf. Section 3.6) did not show differences in age ($BF_{10} = 0.52 \pm 0\%$, $r_{scale} = 0.3$). Overall, affinity for technology interaction was on medium to high level and not differing between groups ($M_{control} = 3.79$, $M_{feedback} = 3.88$, $BF_{10} = 0.78 \pm 0\%$, $r_{scale} = 0.3$, scale range 1–6). The majority of participants stated that they had no vocational qualification or bachelor’s degree. These individuals were probably students at the time of the study. The remaining participants stated to they had a vocational qualification or technician/foreman’s degree (11%) or master’s degree (17%). Therefore, effects of participant groups (students vs. non-students) for baseline trust as the most important outcome of this study was tested. Bayes factors favored the null hypothesis in the overall sample ($BF_{10} = 0.55 \pm 0\%$, $r_{scale} = 0.3$). Slightly more participants of the feedback group (29%) had interacted with an industrial robot before (vs. 20%) and were currently or had ever worked in production sector before (29% vs. 25%). Again, effects of previous personal experience in production (experienced vs. non-experienced) for baseline trust was tested. Bayes factors favored the null hypothesis in the overall sample ($BF_{10} = 2.06 \pm 0\%$, $r_{scale} = 0.3$). Participants received financial compensation.

3.5 Procedures

Participants received the participant information, signed a declaration of consent and filled in the pre-survey. All materials applied to conduct the study including such as experimenter protocols and surveys is public available (cf. Data Availability Statement).

Participants watched two videos showing a real workplace with a handling device, as well as an equivalent task performed collaboratively with the robot in a test environment. This was intended to support the cover story of workplace design in the productive industry. Participants were then given instructions on the collaborative task, which consisted of setting the height via gesture control and carrying out the assembly task (see Table 1, phase 2). Participants in the feedback group also received a brief introduction to using the feedback system. All participants were instructed to:

- Stay within the marked safety zone before the robot stops inside the collaboration zone; leaving the zone would result in an emergency stop,
- Use gesture control for height setting in each assembly cycle,
- Stop the assembly and enter the safety zone as soon as anything abnormal or critical is noticed,
- And minimize their assembly time, as financial compensation would depend on it (compensation manipulation).

All participants performed a baseline condition involving a minimum of ten assembly cycles (see Table 1) in order to familiarize themselves with the assembly task, learn how to control the robot using gestures and become trained with the feedback system and the affective experience assessment. Afterwards, each participant

performed four interaction scenarios with the robot in randomized order. Prior to each condition, participants were asked to rate their current valence and arousal values to establish a baseline (t_0). Valence and arousal values were verbally rated by participants every 10 s (see Section 3.3) following an acoustic signal and recorded by experimenters.

Each experimental scenario comprised five assembly cycles, lasting around 5 min in total. The durations of an assembly cycle resulted from robot periods (35 s) and the individual times taken for gesture control and component assembly. A typical complete assembly cycle without simulated failures lasted 57.5 s ($SD = 4.04$), whereas one containing simulated failures lasted 67.5 s ($SD = 3.97$). System failures were simulated for all participants in the first and third assembly cycle of the respective experimental scenario. The position of simulated failures during the five assembly cycles was not randomized to reduce the mental load for the experimenters, thereby ensuring that all planned and manually simulated system failures were carried out by the experimenters in accordance with the specified procedure. Additionally, the fixed order of the simulated failures occurred only twice for each participant during the experiment; therefore, it can be assumed that the participants were unable to perceive the fixed order.

Each scenario was followed by a short survey to measure outcomes. At the end, participants were debriefed about the financial compensation manipulation and received compensation regardless of their assembly time. Overall, the study lasted about 60 min.

3.6 Data analysis

Statistic Software R (version 4.5.0; R Core Team, 2025) was used for data analysis. Bayes factor analysis was applied, using R package “BayesFactor” (Morey and Rouder, 2024). In accordance with Berger (2006), a very broad prior distribution was used as uninformed prior. It was operationalized by a zero-centered Cauchy distribution with a “medium” prior scale, reflecting common effect sizes in psychological studies (Speekenbrink, 2023). It was implemented by an according rscale parameter of $r = 1/\sqrt{2}$ (Morey and Rouder, 2024) in the `ttestBF()` function to compare independent groups. The parameterization of `anovaBF()` followed the same theoretical considerations and was used to examine main and interaction effects of the three independent variables. Unlike `ttestBF()`, the rscale parameter in `anovaBF()` is based on standardized treatment effects rather than Cohen’s d . Therefore, rscale in `anovaBF()` was set to $r = 0.5$, analogously representing a “medium” prior scale (Speekenbrink, 2023). Individual scores were defined as random effect to address for the within-subjects variables system failure and time pressure. Finally, approximation method “simple” was used to account for the relatively small sample size for complex models (Navarro, 2015). BF_{10} is reported unless otherwise stated and represents the relative likelihood of the alternative hypothesis over the null hypothesis. For the interpretation of Bayes factors, we apply Jeffreys’s scale of evidence (Jeffreys, 1961). With the aim of equivalence testing, in this paper, parameter rscale was set to 0.3 to test the null hypothesis against a rather small effect. For

interpretation, $BF_{10} < 1$ shows a favoring of the null hypothesis over the alternative hypothesis, supporting equivalence of groups or conditions.

Package “ggplot2” (Wickham, 2016) was used to visualize time progression data. Valence and arousal values of each participant were normalized as deviation from initial values queried prior to each scenario starting at t_0 . Additionally, due to individual durations of gesture control and assembly of components, data of phase 2 and 3 (see Table 1) were also normalized using the median values across participants and dependent on scenario. Robot periods also varied as a result of minor technical issues. Therefore, all data within a specified phase were normalized to the means across comparable conditions. To smooth data curves, a generalized additive model (GAM; Wood et al., 2016) was used. According to the relatively low number of data points, the adaption of the GAM model in Wood et al. (2016) was applied and the function parameter “formula = $y \sim s(x, bs = “cs”, fx = T, k = 0)$ ” was used within the function `geom_smooth()` by setting k value to 10 for all graphs aiming at visual comparability. Due to missing robot log files, the resulting diagrams contain a different intersection of the total sample. However, due to the use of relative rather than absolute values and the approximation based on around 900 data points per line, the resulting distortion is expected to be low. Package “lavaan” (Rosseel, 2012) was used for structural equation modeling using a maximum likelihood estimation.

4 Results

4.1 Manipulation check

Due to the two within-subjects variables used, participants experienced two manipulations during the experiment. **System failures** were simulated for about 10 s during the first and third assembly cycle. Consequently, the duration of gesture control should be 10 s longer in experimental conditions involving simulated failures (F+). Figure 4A shows the duration of gesture control depending on the experimental condition, assembly cycle and group. Gesture control was higher in cycles 1 and 3 in experimental conditions with failures (F+; top) than in conditions without simulated failures (F-; bottom), demonstrating the success of the planned manipulation.

Regarding the manipulation of **time pressure**, the available assembly time for each individual (cf. Section 3.2) was reduced. Figure 4B (top) shows the subjective ratings for the NASA-TLX control item “temporal demands.” It is obvious that the manipulation of time pressure was more effective in the feedback group. However, the Bayes factor revealed no significant differences between the groups in either condition involving time pressure and failures (TP+/F+; $BF_{10} = 0.46 \pm 0.01\%$) or condition involving time pressure without simulated failures (TP+/F-; $BF_{10} = 0.39 \pm 0.01\%$). A Bayesian ANOVA applied to the feedback group showed no effect of time pressure ($BF_{10} = 0.26 \pm 1.07\%$) compared to a denominator model using individual scores as a random effect. The control group showed no effect of time pressure, either. Conversely, the objective manipulation check for time pressure (see Figure 4B, bottom), operationalized as the number of components in the queue, supports the success of the manipulation, as participants

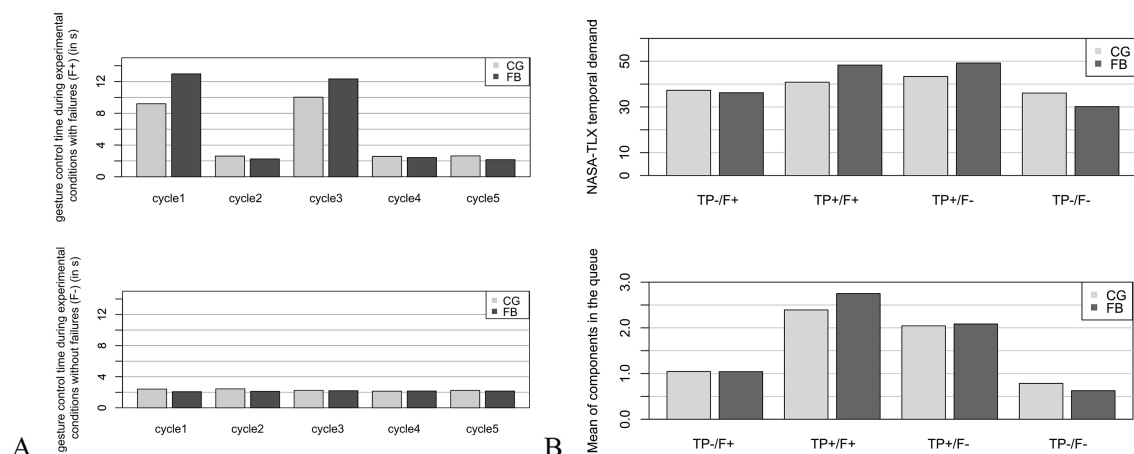


FIGURE 4

Results for manipulation check regarding (A) the realization of system failures operationalized by durations for gesture control dependent on experimental conditions, assembly cycle and group, and (B) time pressure measured as perceived time pressure and objective number of components in the queue at assembly line. CG, control group; FB, feedback group; TP+, with time pressure; TP-, without time pressure; F+, with failures; F-, without failures.

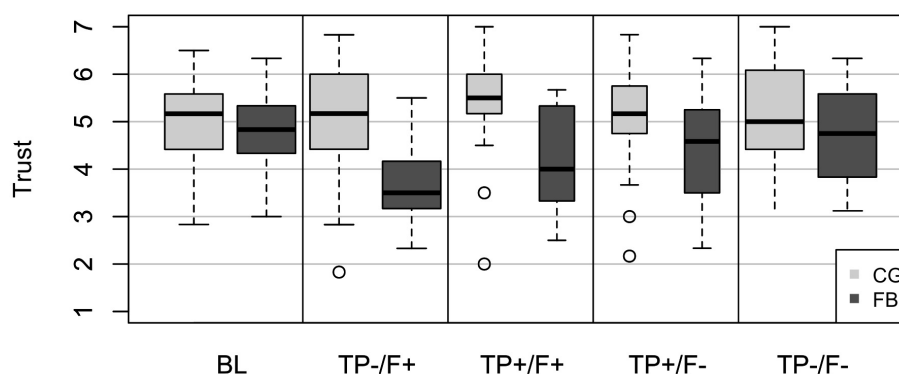


FIGURE 5

Results of trust ratings depending on experimental conditions and group. Small bars show conditions with time pressure (TP+), wide bars without time pressure (TP-). F+ indicates conditions with simulated failure, F- without failure. The lower and upper boundary of the y-axes represents the scale range. BL, baseline; CG, control group; FB, feedback group.

would have been able to reduce their assembly time in order to avoid components in the queue.

4.2 Trust

Figure 5 shows the descriptive results for trust ratings, clustered in group and experimental condition. In the control group, mean trust consistently remained high across all scenarios, whereas trust decreased in the feedback group following a system failure. In the feedback group, trust dropped below the scale mean during experimental condition TP-/F+ ($M = 3.72$, $SD = 0.90$).

Following completion of the baseline condition, trust levels were found to be similar for both groups ($M_{CG} = 4.92$, $M_{FB} = 4.74$, $BF_{10} = 0.54$, $r_{scale} = 0.3$). A Bayesian ANOVA was applied. Table 2 shows the results of selected Bayesian

models compared to a random model, with participants as the random factor. Including only the independent variables “feedback” (row 1) or “system failure” (row 2) significantly improved both models. The effect of “failure” was stronger than of “feedback.” The model including the independent variable “time pressure” (row 3) favored the null hypothesis. Accordingly, the model including the two significant main effects (row 4) showed strong evidence in favor of the alternative hypothesis, and including an interaction of both main effects (row 5) showed an additional significant improvement. Relating Bayesian model results to the descriptive results, (i) overall, trust was lower in the feedback group; (ii) system failures were found to reduce trust, and (iii) system failures had a stronger effect in the feedback group than in the control group. The results support hypothesis 4 (*system failures reduce trust*), but contradict hypotheses 5 (*feedback increases trust*) and 8 (*time pressure increases trust*).

TABLE 2 Bayesian ANOVA results modeling trust compared to a denominator model using participants ID as random factor.

Model	BF ₁₀	Evidence
feedback + ID	6.8 ± 0.73%	Moderate
failure + ID	64.1 ± 0.44%	Very strong
time pressure + ID	0.2 ± 0.63%	(Moderate)
feedback + failure + ID	429.9 ± 0.81	Extreme
feedback + failure + feedback:failure + ID	5175.0 ± 2.17%	Extreme

Colon indicates interaction effects.

4.3 Affective experience

Figure 6 shows how subjective valence and arousal values progressed over time in all four experimental conditions. The valence and arousal values of each participant were normalized as a deviation from the initial values queried prior to the start of each condition, at t_0 . Timelines show the actual *real-time robot periods* (phases 1, 4, and 5), all normalized within the periods across conditions for the purpose of comparison, as well as the normalized data for *gesture control* (phase 2; visualized in green or orange in the event of simulated system failures) and *assembly* (phase 3; visualized in gray) (cf. Section 3.6). A decrease in valence and a simultaneous increase in arousal indicate a negative affective experience (cf. Section 3.3). Changes in valence and arousal that indicate negative affective experience and occur at the same time result in an elliptical shape formed by the lines for valence and arousal.

Comparing all subfigures reveals that conditions involving simulated *failures* (Figures 6A, B) resulted in stronger fluctuations in valence and arousal values than conditions without simulated failures (Figures 6C, D). Conversely, valence and arousal values were not influenced by *time pressure*, as can be seen by comparing Figures 6C, D. A detailed analysis of the conditions involving simulated *failures* (Figures 6A, B) shows that the expected elliptical form of the valence and arousal lines mainly occurred during the assembly cycles involving simulated failures (cycles 1 and 3). This form was more evident in the feedback group than in the control group. Nevertheless, both groups showed a trend toward decreased valence and increased arousal in the first part of each assembly cycle, resulting in maximum deviation in the gesture control or assembly phase. At the end of each assembly cycle (phases 4 and 5), arousal values drop again.

To allow for further quantitative comparisons between the experimental conditions, maximum changes in valence and arousal relative to the preceding value over time were calculated, and the frequency of occurrence depending on the phase was analyzed. Of particular interest to the research questions are negative valence deviations (maximum change hereafter referred to as “*valence drop*”) and positive arousal deviations (maximum change hereafter

referred to as “*arousal surge*”), as these indicate negative affective experiences.

Table 3 shows the relative frequencies of arousal surges depending on the experimental condition and, where applicable, the simulation of system failures in the assembly cycles. Frequency patterns can be identified. Overall, it can be seen from column 9 (“mean”) that arousal surges typically occurred with the following frequency distribution: *assembly* > *gesture control* > *delivery*. Only a few arousal surges occurred during *storage* and *return*. Furthermore, arousal surges occurred more frequently during *gesture control* in cycles with simulated system failures than in cycles without failures, in both experimental groups (cf. columns 3 and 4, and 5 and 6, respectively). However, greater differences were observed in the feedback group. In experimental conditions without failure simulation (cf. columns 7 and 8), arousal surges were rarely caused by *gesture control*, but were more frequently observed in *assembly* (cf. Section 5.3 for a discussion of this result). Phase *return* did not frequently result in arousal surges, even though the robot approached the human at a speed of 500 mm/s while they waited within the safety zone (see Figures 2A, B).

Table 4 shows the relative frequencies of valence drops depending on experimental conditions and, where applicable, the simulation of system failures during assembly cycles. Again, a pattern of frequencies can be identified. Overall, referring to column 9 (“mean”), it is apparent that valence drops typically occurred with the following frequency distribution: *assembly* > *return* > *gesture control*. Only a few valence drops occurred during *delivery*. Furthermore, in both experimental groups, valence drops during *gesture control* were much more frequent in cycles with simulated system failures than in cycles without (cf. columns 3 and 4, and 5 and 6, respectively), and there were no differences between the groups during *gesture control*. In the absence of failure simulations (cf. columns 7 and 8), *gesture control* rarely caused valence drops, but more valence drops were observed during *assembly* (see Section 5.3 for discussion of this result). *Return* caused a high frequency of valence drops in both experimental groups in cycles and conditions without failures comparable to the frequencies found during *assembly*.

To further analyze the occurrence of negative affective experience, the maximum negative deviations of valence and the maximum positive deviations of arousal relative to the initial values (t_0) were calculated. Overall, the maximum valence and arousal changes remained relatively small across all robot phases (compare also Figures 6A–D). A Bayesian $2 \times 2 \times 2$ mixed ANOVA was used to model the maximum deviation in valence and arousal values. Table 5 shows the results of selected Bayesian models compared to a random model, with participants as the random factor. When modeling both valence and arousal changes, it was found that only the model including the independent variable “failure” (row 2) showed significant improvement on the random model. As the models that took the other two independent variables into account favored the null hypothesis (rows 1 and 3), no more complex models are reported. The effect of “failure” was stronger for valence changes than for arousal changes. This result shows that arousal values

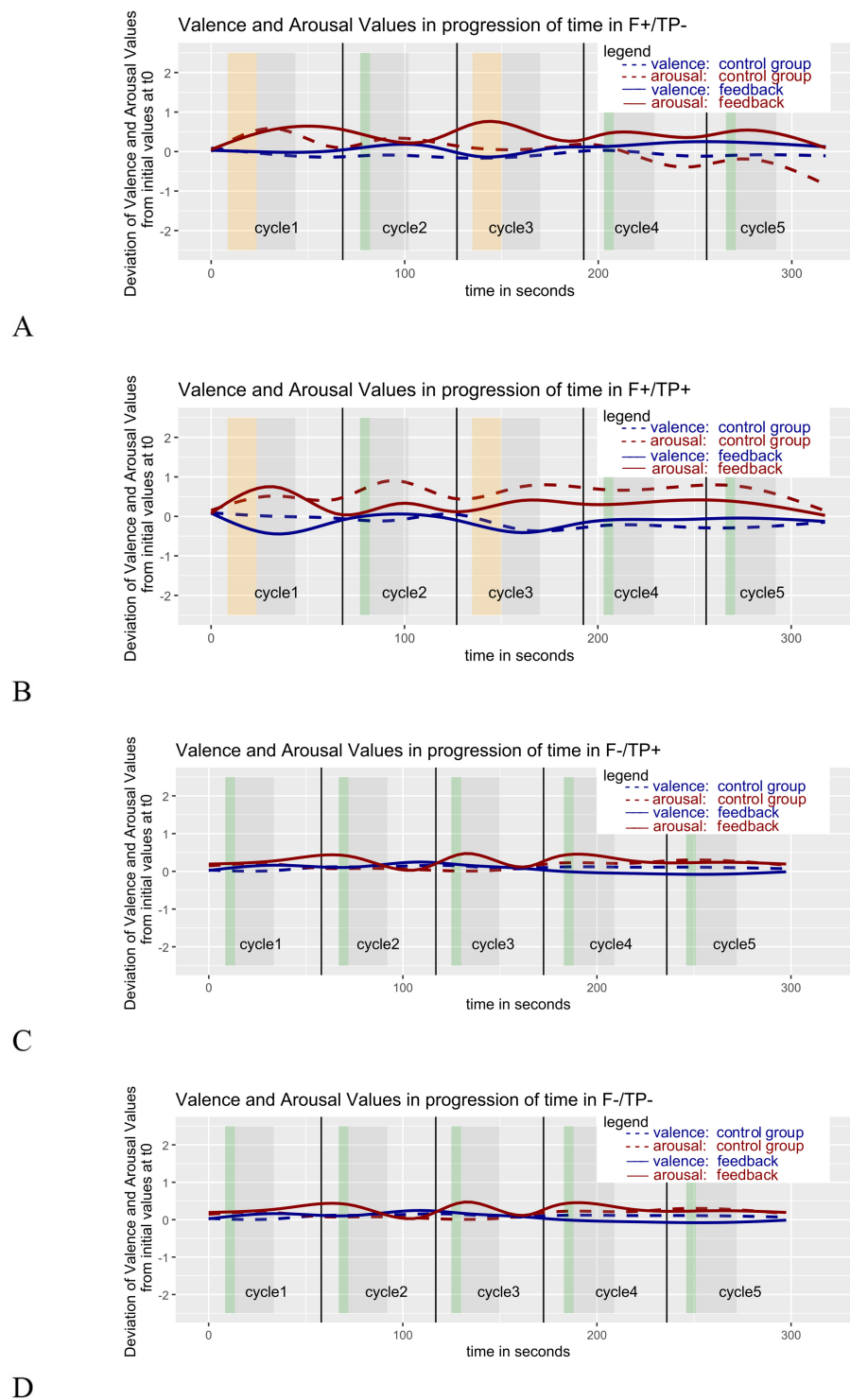


FIGURE 6

Time progression of valence and arousal values depending on experimental condition and group. Duration of gesture control (phase 2) is highlighted in green in conditions without simulated failures and highlighted in orange in case of simulated failures. Assembly time (robot phase 3) is highlighted in gray. t_0 = initial values queried prior to each experimental condition. Graphs belong to condition (A) with simulated failure (F+) but without time pressure (TP-), (B) with simulated failure (F+) and time pressure (TP+), (C) without simulated failures (F-) but time pressure (TP+), and (D) with neither simulated failures (F-) nor time pressure (TP-).

change to some extent also in conditions without system failures, while valence values mostly remain constant (see Figures 6C, D). As reduced valence and increased arousal were found

following system failures, the exploratory hypothesis that system failures induce negative affective experience is supported by the data.

TABLE 3 Relative frequency of arousal surges depending on experimental condition, assembly cycles and group.

Phase	Group	TP-/F+		TP+/F+		TP+/F-	TP-/F-	Mean
		Cycles with failures (1+3)	Cycles without failures (2, 4, 5)	Cycles with failures (1+3)	Cycles without failures (2, 4, 5)	Cycles 1 to 5	Cycles 1 to 5	
Delivery (phase 1)	CG	11%	25%	13%	29%	31%	36%	24%
	FB	5%	15%	4%	20%	24%	33%	17%
Gesture control (phase 2)	CG	35%	6%	43%	19%	7%	7%	20%
	FB	63%	15%	48%	20%	15%	9%	28%
Assembly (phase 3)	CG	32%	53%	30%	45%	42%	44%	41%
	FB	11%	45%	28%	50%	47%	44%	38%
Storage (phase 4)	CG	4%	0%	7%	7%	7%	2%	5%
	FB	16%	0%	12%	0%	6%	2%	6%
Return (phase 5)	CG	18%	16%	7%	0%	13%	11%	11%
	FB	5%	25%	8%	10%	8%	12%	11%

CG, control group; FB, feedback group; TP+, with time pressure; TP-, without time pressure; F+, with failures; F-, without failures.

TABLE 4 Relative frequency of valence drops depending on experimental condition, assembly cycles and group.

Phase	Group	TP-/F+		TP+/F+		TP+/F-	TP-/F-	Mean
		Cycles with failures (1+3)	Cycles without failures (2, 4, 5)	Cycles with failures (1+3)	Cycles without failures (2, 4, 5)	Cycles 1 to 5	Cycles 1 to 5	
Delivery (phase 1)	CG	0%	18%	13%	11%	0%	16%	10%
	FB	0%	0%	13%	14%	4%	0%	5%
Gesture control (phase 2)	CG	54%	0%	27%	0%	5%	5%	15%
	FB	59%	8%	35%	9%	4%	0%	19%
Assembly (phase 3)	CG	38%	18%	33%	33%	53%	37%	35%
	FB	29%	31%	24%	31%	24%	46%	31%
Storage (phase 4)	CG	8%	18%	20%	22%	10%	21%	17%
	FB	6%	15%	13%	10%	24%	27%	16%
Return (phase 5)	CG	0%	46%	7%	34%	32%	21%	23%
	FB	6%	46%	15%	36%	44%	27%	29%

CG, control group; FB, feedback group; TP+, with time pressure; TP-, without time pressure; F+, with failures; F-, without failures.

4.4 Intention to use

Figure 7 shows the descriptive results for the ratings of intention to use, depending on group and experimental condition. The mean intention to use score was consistently higher than the scale mean, showing comparable results across experimental variations and group. Additionally, high variance was found in all conditions and in both groups.

Following completion of the baseline condition, intention to use was found to be similar for both groups ($M_{CG} = 3.64$, $M_{FB} = 3.60$, $BF_{10} = 0.52$, $rscale = 0.3$). Again, a Bayesian ANOVA was applied. Table 6 shows results for selected Bayesian models when compared to a random model with participants as random factor. None of the models showed Bayes factors in favor of the alternative.

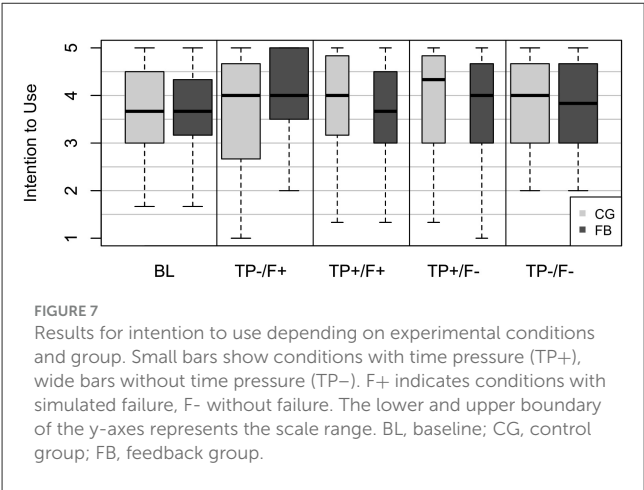
Therefore, the model that only includes participants as a random factor best fits the data.

4.5 Safety-critical behavior

Safety-critical behavior was defined as disregarding simulated system failures during assembly cycles 1 and 3 (cf. Section 3.3). The duration of gesture control during assembly cycles involving system failures was higher for the feedback group ($Mdn = 12.84$, $SD = 3.52$) than for the control group ($Mdn = 9.62$, $SD = 5.49$), with Bayes factor showing different durations of gesture control between groups ($BF_{10} = 17.85 \pm 0\%$) (strong evidence). Figure 8A shows in

TABLE 5 Bayesian ANOVA results modeling maximum negative valence changes and maximum positive arousal changes relative to t_0 compared to a denominator model using participants ID as random factor.

Model	Maximum negative valence changes (relative to t_0)		Maximum positive arousal changes (relative to t_0)	
	BF_{10}	Evidence	BF_{10}	Evidence
Feedback + ID	$0.4 \pm 0.44\%$	(Anecdotal)	$0.38 \pm 0.23\%$	(Anecdotal)
Failure + ID	$11731.8 \pm 0.67\%$	Extreme	$25.9 \pm 0.58\%$	Strong
Time pressure + ID	$0.20 \pm 1.13\%$	(Moderate)	$0.22 \pm 0.73\%$	(Moderate)



more detail that, in the feedback group, durations rarely fell below the defined cut-off value of 8.5 s for “disregarding system failures.” No differences were observed in the disregard of simulated system failures dependent on the time pressure condition. Therefore, when both experimental conditions involving simulated system failures were combined, the control group disregarded 47% of all simulated system failures in assembly cycle 1 and 42% in assembly cycle 3. By contrast, the feedback group disregarded only 5% of failures in assembly cycle 1 and 3% in assembly cycle 3. Bayes factor analysis showed extreme evidence for a difference of both groups ($BF_{10} = 14367.0 \pm 0\%$) (see Figure 8B). As can be seen in Figure 8B, comparing assembly cycles 1 and 3, the frequency of disregarding system failure decreased sparsely.

4.6 Effects of trust: relationship of trust and human-centered outcome criteria

To analyze the relationship between trust and the human-centered outcomes, the experimental condition with the greatest variance in trust ratings was selected. Therefore, non-parametric correlations were calculated for the control group (cf. Table 7) and the feedback group (cf. Table 8) for all dependent variables using experimental condition TP+/F+. In the control group, trust shows a weak positive correlation with safety-critical behavior. The data support hypothesis 3, and not surprisingly, this correlation disappears in the feedback group because hardly any safety-critical behavior was observed in this group (see Section 4.5).

TABLE 6 Bayesian ANOVA results modeling intention to use compared to a denominator model using participants ID as random factor.

Model	BF_{10}	Evidence
Feedback + ID	$0.48 \pm 0.84\%$	(Anecdotal)
Failure + ID	$0.16 \pm 1.22\%$	(Moderate)
Time pressure + ID	$0.52 \pm 0.46\%$	(Anecdotal)
Group + feedback + time pressure + ID	$0.04 \pm 1.29\%$	(Strong evidence)

In the control group, a strong positive correlation was found between trust and intention to use (data supporting hypothesis 2). Again, this correlation decreases in the feedback group, although a weak positive correlation is still evident. While the correlations show the expected direction for arousal surge, trust was non-significantly related to valence drops and arousal surge (data contradict hypothesis 1).

To uncover any potential suppressing effects, a structural equation model was calculated using all variables as manifest variables to estimate the models with trust as a mediator. As safety-critical behavior was only operationalized in experimental conditions involving simulated failures, and as time pressure had no effect on the human-centered outcome criteria (contradicting hypothesis 9, *trust mediates the relationship between time pressure and intention to use, as well as safety-critical behavior*), separate models were calculated for the experimental conditions involving system failures. These models included the independent variable “feedback” as an exogenous variable, trust as a mediator, and valence drop, arousal surge, safety-critical behavior, and intention to use as endogenous variables. All variables were z-standardized to obtain standardized beta weights.

Figure 9 shows the model for the two experimental conditions involving simulated system failures. In condition TP-/F+, the model fit improved significantly when a direct path was added between feedback and safety-critical behavior ($CFI = 0.908$, $RMSEA = 0.156$, $SRMR = 0.075$) (Figure 9A). The same was true for the model for condition TP+/F+ ($CFI = 0.994$, $RMSEA = 0.038$, $SRMR = 0.057$) (Figure 9B). Results show a strong direct, negative effect of feedback on safety-critical behavior (supporting hypothesis 7—*feedback shows an independent effect on reducing safety-critical behavior*) and a strong direct, negative effect of feedback on trust (contradicting hypothesis 5—*feedback increases trust*). Besides, the model for condition TP+/F+ shows a significant mediation effect of trust on safety-critical behavior as the bootstrap confidence interval (using Bollen-Stine bootstrap

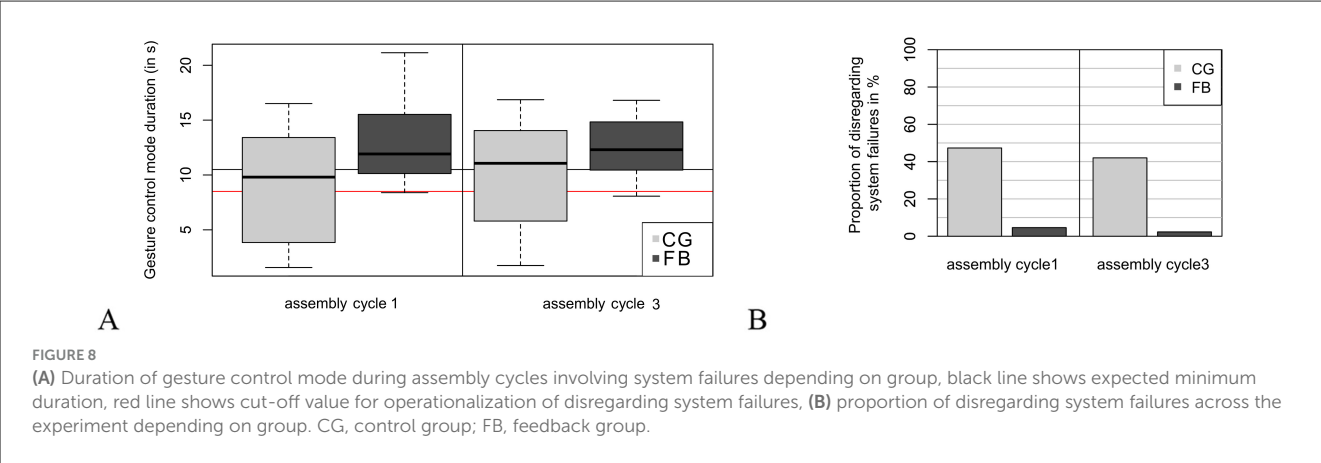


TABLE 7 Non-parametric correlation matrix (Kendall's τ) of all dependent variables in the control group (* $p < 0.05$).

	1	2	3	4	5
1 Trust	–				
2 Valence drops	0.080	–			
3 Arousal surges	–0.153	0.297	–		
4 Safety-critical behavior (disregarding failures)	0.280	<0.001	–0.024	–	
5 Intention to use	0.575*	0.184	0.028	0.193	–

TABLE 8 Non-parametric correlation matrix (Kendall's τ) of all dependent variables in the feedback group.

	1	2	3	4	5
1 Trust	–				
2 Valence drops	0.073	–			
3 Arousal surges	–0.202	–0.151	–		
4 Safety-critical behavior (disregarding failures)	–0.011	–0.026	<0.001	–	
5 Intention to use	0.240	0.006	–0.146	–0.139	–

method, Kim and Millsap, 2014) for estimating the indirect effect is not including zero (CI = [–0.554; –0.014], cf. Table 9). Data support hypothesis 6 (*trust mediates the relationship of feedback with safety-critical behavior*). Data do not support this mediation effect in experimental condition TP–/F+ (CI = [–0.555; 0.274]) (data contradicting hypothesis 6). For both experimental conditions involving simulated failures, the models support the significant mediation effect of trust on intention to use (data support Hypothesis 6—*trust mediates the relationship between feedback and intention to use*). However, there were no mediation effects of trust on valence drop or arousal surge.

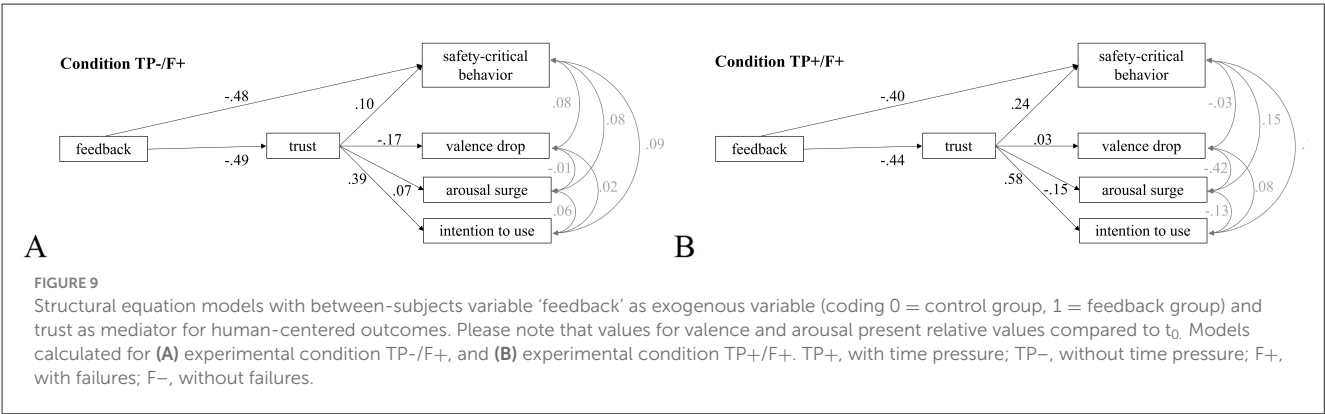
5 Discussion

5.1 Summary and interpretation of results

This empirical study aimed to understand how changes in trust, caused by antecedents of trust (system failures, visual feedback and time pressure) impact human-centered outcomes (affective experience, intention to use and safety-critical behavior). The results showed most surprisingly that time pressure had no effect on trust or any other human-centered outcome. On the other hand, both system failures and the availability of feedback, as expected, were shown to have a significant effect on **trust**. However, the availability of feedback did not affect trust in the absence of system failures. Providing more information to participants about the current and future state of the robot or the time available for assembly did not have a positive effect on trust (cf. Figure 1;

cf. Legler et al., 2022 for details on the feedback system). On the contrary, feedback had a negative effect on trust after the occurrence of system failure—a result that was the opposite of that predicted by our literature-derived hypothesis. This probably indicates that the feedback system focused participants' attention (Goldstein, 2010) on the system failure. For safety reasons, the failure was simulated as a technical malfunction in gesture control, ensuring that participants would not be physically harmed if they disregarded the failure. The reduction in trust observed in the feedback group demonstrates that even a less salient and harmless simulated system failure can significantly diminish trust in robots. As trust remained consistently high in the control group across conditions, the results suggest that the level of trust exceeded an appropriate trust level—a state known as overtrust (Lee and See, 2004). As the participants had no secondary task, they were able to focus exclusively on the robot. As they were also not permitted to assemble in the event of any robot movement, it can be assumed that the participants were able to perceive the malfunction but chose not to react. Such complacency is a typical indicator of overtrust (Parasuraman and Manzey, 2010). Therefore, data leads us to suppose that workers overtrusting a robot might notice system failures, but as a result of their inappropriately high level of trust they choose not to react (cf. "free choice," Schäfer et al., 2024) as they do not expect the robot to work imperfectly or being harmful.

System failures were also found to affect the occurrence of negative affective experiences, measured by valence drops and arousal surges. These negative affective experiences occurred during specific HRC phases. In the case of an experimental



condition involving simulated failures, these phases were *gesture control* and *assembly* in both groups (see [Tables 3, 4](#)). Results show that the negative affective experiences were caused by the simulated system failures. This further supports the assumption that participants in the control group also noticed the system failures, as they showed equivalent changes in valence and arousal compared to the feedback group. Overall, changes in valence and arousal values were within a range of ± 2 points on an 11- or 12-point scale, respectively. Although these time-synchronized deviations in both values indicate a negative affective experience during system failures, the intensity of the resulting affects is assumed to be within a range that does not cause distress to humans.

Regarding intention to use, no effect of any of the independent variables was found in the data. Scores remained high across conditions and showed high variance due to individual dispositions, assessed by high intraindividual consistency of intention to use. However, intention to use did not correlate with negative affective experience (see [Tables 7, 8](#)), indicating that low intention to use was not caused by perceived distress in participants, but rather by an individual tendency to reject (new) technology in the workplace.

Safety-critical behavior was rarely observed in the feedback group, but frequently in the control group. This result again highlights the importance of feedback in HRC. The negative correlation of trust and safety-critical behavior has only been found in the control group, demonstrating that feedback has the potential to counteract individual tendencies toward overtrust.

Overall, the effect sizes of the relationships between trust, affective experience, intention to use, and safety-critical behavior demonstrate the distinctiveness of all human factors constructs applied. Trust was found to be unrelated to self-reported changes in valence and arousal. There are several possible explanations for this result. Firstly, changes in affective experience were assessed during the interaction, whereas trust was measured after the interaction. As emotions are fleeting ([Werth and Förster, 2015](#)), this result may have been confounded by this methodological choice. Secondly, the changes in valence and arousal may have been too small to detect a significant correlation. Third, from a theoretical perspective, the results could suggest that trust is a concept with stronger cognitive than affective aspects, as evidenced by the variety of definitions of trust in robots (cf. [Schaefer, 2013](#) for a detailed overview),

TABLE 9 Bootstrap confidence interval for the mediation effect of “trust” in two experimental conditions involving simulated system failures.

Mediation effect	CI (Bollen-Stine bootstrapping method)	
	TP-/F+	TP+/F+
Feedback → trust → safety-critical behavior	[−0.555; 0.274]	[−0.554; −0.014]
Feedback → trust → valence drop	[−0.108; 0.654]	[−0.321; 0.247]
Feedback → trust → arousal surge	[−0.617; 0.251]	[−0.105; 0.415]
Feedback → trust → intention to use	[−0.750; −0.061]	[−0.973; −0.216]

which conceptualize trust as an “expectation,” “belief” or “attitude.” This could also explain the strong correlation between trust and intention to use. Beyond trust, the other human-centered outcomes showed no significant correlations with each other. These results further support a need to integrate varieties of human factors criteria when studying HRC in the advent of Industry 5.0.

Finally, it was demonstrated that trust in robots mediated the effect of human-centered outcomes in industry, i.e., safety-critical behavior and intention to use. As strategies for intentionally adjusting trust levels have already been studied ([de Visser et al., 2020](#)), our results show that such strategies have a potential to affect industry-relevant outcomes such as safety-critical behavior—thus going beyond adjusting trust levels, which has unclear effects on HRC. This further highlights the importance of considering trust in HRC research to realize the human-centered vision of Industry 5.0.

5.2 Practical implications for HRC design in industrial workplaces

Several implications for industrial HRC design can be derived from the presented study results.

1. Feedback is highly relevant in ensuring situation awareness during HRC. It enables humans to perceive system failures and react properly to these off-nominal, infrequent and unexpected situations ([Hopko et al., 2022](#)). As malfunctions of robots are

infrequent but always possible (Schäfer et al., 2024) and taking high costs for manufacturing companies both in case of health impairments of employees and reduced quality of products into account, high situation awareness of humans in HRC is key to success. Although this was not assessed in the study, situation awareness is especially important in the case of parallel human tasks (“dual tasks”), which are expected to be the norm in industrial HRC workplaces in order to make effective use of the autonomous robot periods. Nevertheless, dual tasks prevent humans from constantly monitoring the robot’s actions, which again underlines the importance of feedback.

2. Simulating a set of safe system failures that do not negatively impact performance outcomes could maintain trust at an “appropriate level” by encouraging workers to pay attention to and monitor the behavior of the robot. This reduces the potential risk of overtrust arising from longer interactions with failure-free automation. These simulated failures must be made salient using a feedback system to reduce the risk of them being missed. This proposition aligns with HRI research which suggests familiarizing users with potential robot failures (Wagner et al., 2018) and applying “deceptive practices” of robots (Aroyo et al., 2021) to prevent overtrust.
3. Trust was unrelated to self-reported changes in valence and arousal. This calls into question frequently suggested operationalizations of trust measurement based on physiological arousal (Arai et al., 2010; Hopko et al., 2022). Additionally, physiological measures require a person to be wired with sensors, which raises questions not only of comfort, but also of protecting personal data. Instead, aiming at a recognition of human trust by an automated robotic system in the future, behavioral indicators of trust (Chang and Hasanzadeh, 2024) should preferably be used, as these are non-invasive and do not distract workers from their tasks.

5.3 Limitations and future research directions

The study presented had some limitations. From a content perspective, as the study focused on *human-centered outcomes*, performance-centered outcomes were not considered. While recognizing the importance of human-centered outcomes, performance-centered outcomes remain an indispensable target variable in Industry 5.0, as performance is key to success in manufacturing. Investigations into assembly time as a performance-centered outcome can be found in Legler et al. (2022). No effects on assembly time were found as a result of industry-relevant antecedents of trust. There are two possible explanations for this effect. Firstly, even under time pressure, the assembly task may not have been complex enough to affect objective performance measurements when performed without a secondary task. Secondly, it might not be possible to detect human adaptation of assembly time within a total *duration* of 5 min per scenario. As also trust is known to develop over a longer period of time, studies involving longer, undisturbed interaction with the robot are required. Nevertheless, high levels of trust

were already evident in baseline conditions, indicating positive expectations toward robots prior to the interaction. Furthermore, in real industrial scenarios, trust is equally expected to increase during failure-free HRC.

Due to the post-scenario measurement and laboratory setting, it can be assumed that participants were aware of occurring experimental variations and anticipated some form of manipulation. Flook et al. (2019) summarized that ecological validity is low in *laboratory settings*, especially in the case of failure simulation, as participants perceive the setting as artificial, controlled, and therefore safe. However, it can be assumed that workers in industrial HRC workplaces also believe their workplaces to be safe due to occupational safety examinations prior to workplace release. *Time pressure* had no effect on the human-centered outcomes, equally calling into question the ecological validity, as the participants probably did not expect consequences if demonstrating insufficient performance.

Finally, this study cannot imply the *long-term* effects of implementing safe system failures to adjust trust toward a more appropriate level, nor the possible habituation to “safe” system failures of workers at real HRC workplaces. Although participants were instructed to assemble only at the component on the robot flange when the robot stopped moving, participants in the control group may have perceived the simulated *system failures* but judged them as non-critical and therefore disregarded them. Behavioral observations during the experiment and correct gesture control during the baseline support this explanation. Still, the instruction of participants prior to interaction combined with the baseline condition in this study is comparable to workplace trainings that often include verbal or written instructions for behavior in exceptional situations while practical training is only carried out for normal work processes.

As trust and affective experience has been shown to be unrelated in this study, the findings show a need for a theoretical framework that can explain how and under which circumstances trust impacts emotions.

Several methodological limitations due to the laboratory study design also need to be discussed in order to interpret the results.

Given the relatively small *sample* of the study, a Bayes factor analysis was performed to compare the relative evidence for the null compared to the alternative hypothesis. Overall, the effects observed in the study clearly favored either the null or the alternative hypothesis. However, the relatively small sample size could account for some effects in this study classified as anecdotal evidence (cf. Jeffreys, 1961) which means that the data show a particular degree of uncertainty regarding the favoring of the null or alternative hypothesis. These effects should be investigated in further studies. Additionally, the sample did not consist of actual workers from the productive industry. Based on participants’ highest educational degree, it can be assumed that more than two-thirds of participants were students at the time of the study. Although no statistical effects were found when comparing students and non-students, or participants with and without prior production experience, the generalizability of the results has to be judged as limited, and the results of the study need to be replicated with actual workers from the productive industry. As a starting point, a sample of experienced workers

from workplaces including fenceless human-robot coexistence (no shared work task) seem promising.

The *cut-off value* used to operationalize the disregard of system failures has a significant impact on the results regarding safety-critical behavior. Nevertheless, the difference between groups in terms of disregarding system failures was large enough for the effect to remain when the defined tolerance was deducted.

Regarding the *assessment of affective experience* over time, the sampling rate of valence and arousal values at 10-s intervals reveals that negative affective experiences probably occurred during a different HRC phase to the one referred to during data analysis. For example, a surge in arousal could have occurred during *gesture control*, but been reported at the next assessment point in *assembly*. Nevertheless, trends across the different phases of the robot can be inferred from the data and pretest results on assessment intervals (cf. Legler et al., 2023) showed that a higher sampling rate would have reduced the focus on the primary collaboration task. Still, it cannot be completely ruled out that the chosen interval of assessment distracted participants or that they reduced the additional mental workload of self-report by anchoring themselves to their previous values. To address these uncertainties, relative values were evaluated instead of absolute valence and arousal values and values were only interpreted in group or condition comparison as the assessment was applied equally throughout the entire study.

The current study relied exclusively on *self-report measures* to assess affective experience, trust, and intention to use. Using objective measures would allow robotic systems to autonomously detect changes in human experience and respond by adapting their behaviors, e.g., by adjusting their path planning. However, as physiological measurements are non-specific (Mauss and Robinson, 2009), invasive and are prone to motion artifacts in situations involving a lot of movement (Chang and Hasanzadeh, 2024), such as an assembly task, future research should investigate how human factors constructs like trust can be validly measured over time. This would also provide further insight into changes in trust over time during HRC. Particular attention should be given to behavioral measurements of trust, as these remain understudied (Chang and Hasanzadeh, 2024). As trust has been found to be negatively related to monitoring behavior (Hergeth et al., 2016), the use of non-intrusive eye-tracking (e.g., Pluchino et al., 2023) devices during HRC or even a robotic detection of human gaze based on optical face recognition (e.g., Al Naser et al., 2022) seems promising for online trust assessment.

Results of our study suggest that overtrust in robots may occur in HRC. However, there are currently no specific criteria to determine an “appropriate trust level” that could be used to decide whether trust repair or trust dampening should be applied (de Visser et al., 2020). To enable real-time adjustment of human trust during HRC, there is a pressing need for research that quantitatively explore appropriate trust, in order to overcome the current theoretical considerations and wording of “avoiding under- and overtrust.” Furthermore, the present study highlights the need for further research into the consequences of trust, examining the relationship between trust and outcomes relevant to ensuring safety and efficiency of HRC.

6 Conclusion

With the aim of achieving human-centered automation, this paper investigates the largely neglected relationship of trust in robots with industry-relevant affective, cognitive, and behavioral consequences in industrial human–robot collaboration (HRC). To this end, the paper reviews the current state of the art of trust in HRC, putting trust at the center of attention. The paper highlights the relationship between trust and affective experience, intention to use, and safety-critical behavior, extending it to potential mediating effects of trust between industrial-relevant antecedents of trust and these human-centered outcomes, from a theoretical perspective. The antecedents of trust and the outcomes are then linked to the key characteristics of a recent, comprehensive definition of trust in robots. An empirical contribution to our understanding of the role of trust in industrial HRC is made by conducting an experimental, scenario-based study. By varying trust-related characteristics of industrial HRC—system failures, feedback and time pressure—in a pseudo real-world assembly scenario, it was demonstrated that these characteristics identified in trust literature have the potential to affect intention to use and safety-critical behavior during HRC. Combining system failures with feedback significantly reduced high levels of trust to a still tolerable level without causing distress to participants, but significantly increased proper reactions to failure events. Without feedback, trust consistently remained at a high level even after system failures occurred, indicating a lack of awareness that could potentially enhance reaction times in safety-critical situations. The paper emphasizes the importance of investigating the mediating effects of trust during HRC on industry-relevant outcomes, especially human-centered outcomes, and provides implications for practical HRC design and directions for future research.

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 below: https://figshare.com/projects/Legler_Bullinger_Front_Org_Psych_2025/256862.

Ethics statement

The studies involving humans were approved by the Ethical Commission of the Faculty of Social and Behavioral Sciences at FSU Jena (process number FSV 18/26). 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.

Author contributions

FL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AB: Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research took place within the scope of project “3DIMiR” (funding code 03ZZ0459D) supported by the Federal Ministry of Research, Technology and Space. The authors acknowledge this financial support. The funding agency did not have an impact on study design, data acquisition, analysis and interpretation of data as well as authoring and submission of this article. The publication of this article was funded by Chemnitz University of Technology.

Acknowledgments

We are grateful to our project partners Dr. Mohamad Bdiwi, Lena Winkler, and Shuxiao Hou from the Department of Cognitive Human-Machine-Systems at Fraunhofer IWU for the realization of the test environment. Furthermore, we thank the Ethical Commission of the Faculty of Social and Behavioral Sciences at FSU Jena (institution of project partners) for ethical approval. We are grateful to Selina Palige for her assistance in data preparation. Finally, we are grateful to the Human Factors and Ergonomics Society Europe Chapter for giving us the opportunity to present our research.

References

- Akalin, N., Kiselev, A., Kristoffersson, A., and Loutfi, A. (2023). A taxonomy of factors influencing perceived safety in human-robot interaction. *Int. J. Soc. Robot.* 15, 1993–2004. doi: 10.1007/s12369-023-01027-8
- Al Naser, I., Dahmen, J., Bdiwi, M., and Ihlenfeldt, S. (2022). “Fusion of depth, color, and thermal images towards digital twins and safe human interaction with a robot in an industrial environment,” in *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Napoli: IEEE), 532–537. doi: 10.1109/RO-MAN53752.2022.9900548
- Alarcon, G. M., Lyons, J. B., Hamdan, I., aldin, and Jessup, S. A. (2024). Affective responses to trust violations in a human-autonomy teaming context: humans versus robots. *Int. J. Soc. Robot.* 16, 23–35. doi: 10.1007/s12369-023-01017-w
- Alves, J., Lima, T. M., and Gaspar, P. D. (2023). Is industry 5.0 a human-centred approach? A systematic review. *Processes* 11:193. doi: 10.3390/pr11010193
- Arai, T., Kato, R., and Fujita, M. (2010). Assessment of operator stress induced by robot collaboration in assembly. *CIRP Ann.* 59, 5–8. doi: 10.1016/j.cirp.2010.03.043
- Aroyo, A. M., de Bruyne, J., Dheu, O., Fosch-Villaronga, E., Gudkov, A., Hoch, H., et al. (2021). Overtrusting robots: setting a research agenda to mitigate overtrust in automation. *Paladyn. J. Beh. Robot.* 12, 423–436. doi: 10.1515/pjbr-2021-0029
- Babamiri, M., Heidarimoghdam, R., Ghasemi, F., Tapak, L., and Mortezaipoor, A. (2022). Insights into the relationship between usability and willingness to use a robot in the future workplaces: studying the mediating role of trust and the moderating roles of age and STARA. *PLoS ONE* 17:e0268942. doi: 10.1371/journal.pone.0268942
- Bdiwi, M., Krusche, S., and Putz, M. (2017a). “Zone-based robot control for safe and efficient interaction between human and industrial robots,” in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, eds. B. Mutlu, M. Tscheligi, A. Weiss, and J. E. Young (New York, NY: ACM), 83–84. doi: 10.1145/3029798.3038413
- Bdiwi, M., Pfeifer, M., and Sterzing, A. (2017b). A new strategy for ensuring human safety during various levels of interaction with industrial robots. *CIRP Ann.* 66, 453–456. doi: 10.1016/j.cirp.2017.04.009
- Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Anal.* 1, 385–402. doi: 10.1214/06-BA115
- Brauer, R. R. (2017). *Akzeptanz kooperativer Roboter im industriellen Kontext [Acceptance of cooperative robots in the industrial context]* (dissertation). Universitätsverlag Chemnitz, Chemnitz. Available online at: <https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa-228852>
- Buxbaum, H.-J., and Sen, S. (2021). “Soziotechnische Robotersysteme in Produktion, Pflege und Alltag – neue Wege der Mensch-Maschine-Interaktion: ein Ansatz für neue Interaktionsmöglichkeiten zwischen Mensch und Maschine durch Kombination der industriellen Mensch-Roboter-Kollaboration mit der sozialen Robotik,” in *Soziale Roboter*, ed. O. Bendel (Wiesbaden: Springer Fachmedien Wiesbaden), 379–397. doi: 10.1007/978-3-658-31114-8_20
- Chang, W.-C., and Hasanzadeh, S. (2024). Toward a framework for trust building between humans and robots in the construction industry: a systematic review of current research and future directions. *J. Comput. Civ. Eng.* 38:03124001. doi: 10.1061/JCCEE5.CPENG-5656
- Coronado, E., Kiyokawa, T., Ricardez, G. A. G., Ramirez-Alpizar, I. G., Venture, G., and Yamanobe, N. (2022). Evaluating quality in human-robot interaction: a systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0. *J. Manuf. Syst.* 63, 392–410. doi: 10.1016/j.jmsy.2022.04.007
- de Visser, E., Peeters, M. M. M., Jung, M., Kohn, S., Shaw, T., Pak, R., et al. (2020). Towards a theory of longitudinal trust calibration in human-robot teams. *Int. J. Soc. Robot.* 12, 459–478. doi: 10.1007/s12369-019-00596-x
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., and Beck, H. P. (2003). The role of trust in automation reliance. *Int. J. Hum.-Comput. Stud.* 58, 697–718. doi: 10.1016/S1071-5819(03)00038-7
- Flook, R., Shrinah, A., Wijnen, L., Eder, K., Melhuish, C., and Lemaignan, S. (2019). On the impact of different types of errors on trust in human-robot interaction: are laboratory-based HRI experiments trustworthy? Interaction studies. *Soc. Behav. Commun. Biol. Artif. Syst.* 20, 455–486. doi: 10.1075/is.18067.FLO
- Franke, T., Attig, C., and Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. *Int. J. Hum.-Comput. Interact.* 35, 456–467. doi: 10.1080/10447318.2018.1456150

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.

Any alternative text (alt text) provided alongside figures in this article 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.

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.

- Giallanza, A., La Scalia, G., Micale, R., and La Fata, C. M. (2024). Occupational health and safety issues in human-robot collaboration: state of the art and open challenges. *Saf. Sci.* 169:106313. doi: 10.1016/j.ssci.2023.106313
- Goldstein, E. B. (2010). *Sensation and Perception, 8th Edn.* Belmont, CA: Wadsworth Cengage Learning.
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Hum. Fact.* 53, 517–527. doi: 10.1177/0018720811417254
- Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., and Szalma, J. L. (2021). Evolving trust in robots: specification through sequential and comparative meta-analyses. *Hum. Fact.* 63, 1196–1229. doi: 10.1177/0018720820922080
- Hardy, C. J., and Rejeski, W. J. (1989). Not what, but how one feels: the measurement of affect during exercise. *J. Sport Exerc. Psychol.* 11, 304–317. doi: 10.1123/jsep.11.3.304
- Hart, S. G. (2006). Nasa-task load index (NASA-TLX); 20 years later. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 50, 904–908. doi: 10.1177/154193120605000909
- Hergeth, S., Lorenz, L., Vilimek, R., and Krems, J. F. (2016). Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. *Hum. Fact.* 58, 509–519. doi: 10.1177/0018720815625744
- Hopko, S., Wang, J., and Mehta, R. (2022). Human factors considerations and metrics in shared space human-robot collaboration: a systematic review. *Front. Robot. AI* 9:799522. doi: 10.3389/frobt.2022.799522
- Hopko, S. K., Mehta, R. K., and Pagilla, P. R. (2023). Physiological and perceptual consequences of trust in collaborative robots: an empirical investigation of human and robot factors. *Appl. Erg.* 106:103863. doi: 10.1016/j.apergo.2022.103863
- ISO International Organization for Standardization (2011). *ISO 10218-1: Robots and Robotic Devices - Safety Requirements For Industrial Robots - Part 1: Robots.* Geneva, Switzerland: ISO.
- ISO International Organization for Standardization (2016). *ISO/TS 15066:2016: Robots and robotic devices — Collaborative robots.* Geneva, Switzerland: ISO.
- Jeffreys, H. (1961). *Theory of Probability, 3rd Edn.* Oxford: Oxford University Press.
- Khalid, A., and Syed, J. (2024). Mental health and well-being at work: a systematic review of literature and directions for future research. *Hum. Resour. Manag. Rev.* 34:100998. doi: 10.1016/j.hrmr.2023.100998
- Kim, H., and Millsap, R. (2014). Using the bollen-stine bootstrapping method for evaluating approximate fit indices. *Multivariate Behav. Res.* 49, 581–596. doi: 10.1080/00273171.2014.947352
- Kopp, T. (2024). Facets of trust and distrust in collaborative robots at the workplace: towards a multidimensional and relational conceptualisation. *Int. J. Soc. Robot.* 16, 1445–1462. doi: 10.1007/s12369-023-01082-1
- Kopp, T., Baumgartner, M., and Kinkel, S. (2021). Success factors for introducing industrial human-robot interaction in practice: an empirically driven framework. *Int. J. Adv. Manuf. Technol.* 112, 685–704. doi: 10.1007/s00170-020-06398-0
- Lee, J. D., and See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Hum. Fact.* 46, 50–80. doi: 10.1518/hfes.46.1.50.30392
- Legler, F., Langer, D., Dittrich, F., and Bullinger, A. C. (2020). “I don’t care what the robot does! Trust in automation when working with a heavy - load robot,” in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference: Understanding Human Behaviour in Complex Systems*, eds. D. de Waard, A. Toffetti, L. Pietrantonio, T. Franke, J.-F. Petiot, C. Dumas, et al., 239–253. Available online at: <https://www.hfes-europe.org/largefiles/proceedingshfeurope2019.pdf> (Accessed July 20, 2025).
- Legler, F., Langer, D., Glende, S., and Bullinger, A. C. (2022). “Ten seconds to go! – effects of feedback systems in human-robot collaboration,” in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference: Enhancing Safety Critical Performance*, eds. D. de Waard, S.H. Fairclough, K.A. Brookhuis, D. Manzey, L. Onnasch, A. Naumann, et al., 161–177. Available online at: <https://www.hfes-europe.org/largefiles/proceedingshfeurope2022.pdf> (Accessed July 20, 2025).
- Legler, F., Trezl, J., Langer, D., Bernhagen, M., Dettmann, A., and Bullinger, A. C. (2023). Emotional experience in human-robot collaboration: suitability of virtual reality scenarios to study interactions beyond safety restrictions. *Robotics* 12:168. doi: 10.3390/robotics12060168
- Lewis, M., Sycara, K., and Walker, P. (2018). “The Role of Trust in Human-Robot Interaction,” in *Foundations of Trusted Autonomy*, eds. H. A. Abbass, J. Scholz, and D. J. Reid (Cham: Springer International Publishing), 135–159. doi: 10.1007/978-3-319-64816-3_8
- Luhmann, N. (1979). *Trust and Power: Two Works by Niklas Luhmann.* Chichester: Wiley.
- Maurtua, I., Ibarguren, A., Kildal, J., Susperregi, L., and Sierra, B. (2017). Human-robot collaboration in industrial applications. *Int. J. Adv. Robot. Syst.* 14:172988141771601. doi: 10.1177/1729881417716010
- Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cogn. Emot.* 23, 209–237. doi: 10.1080/02699930802204677
- Morey, R. D., and Rouder, J. N. (2024). *BayesFactor: Computation of Bayes Factors for Common Designs.* Available online at: <https://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf>
- Murray, E. G., Neumann, D. L., Moffitt, R. L., and Thomas, P. R. (2016). The effects of the presence of others during a rowing exercise in a virtual reality environment. *Psychol. Sport Exerc.* 22, 328–336. doi: 10.1016/j.psychsport.2015.09.007
- Navarro (2015). *Learning Statistics With R: A Tutorial for Psychology Students and Other Beginners. (Version 0.6.1).* Available online at: <https://learningstatisticswithr.com/book/> (Accessed July 8, 2025).
- Onnasch, L., and Hildebrandt, C. L. (2022). Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction. *J. Hum.-Robot Interact.* 11, 1–24. doi: 10.1145/3472224
- Oubari, A., Pischke, D., Jenny, M., Meißner, A., and Trübsetter, A. (2018). Mensch-Roboter-Kollaboration in der produktion: motivation und einstellungen von entscheidungsträgern in produzierenden unternehmen. *ZWF Z. fuer Wirtsch. Fabr.* 113, 560–564. doi: 10.3139/104.111971
- Paliga, M. (2023). The relationships of human-cobot interaction fluency with job performance and job satisfaction among cobot operators—the moderating role of workload. *Int. J. Environ. Res. Public Health.* 20:5111. doi: 10.3390/ijerph20065111
- Parasuraman, R., and Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Hum. Fact.* 52, 381–410. doi: 10.1177/0018720810376055
- Pietrantonio, L., Favilla, M., Fraboni, F., Mazzoni, E., Morandini, S., Benvenuti, M., et al. (2024). Integrating collaborative robots in manufacturing, logistics, and agriculture: expert perspectives on technical, safety, and human factors. *Front. Robot. AI* 11:1342130. doi: 10.3389/frobt.2024.1342130
- Pluchino, P., Pernice, G. F. A., Nenna, F., Mingardi, M., Bettelli, A., Bacchin, D., et al. (2023). Advanced workstations and collaborative robots: exploiting eye-tracking and cardiac activity indices to unveil senior workers’ mental workload in assembly tasks. *Front. Robot. AI* 10:1275572. doi: 10.3389/frobt.2023.1275572
- Pöhler, G., Heine, T., and Deml, B. (2016). Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen. *Z. Arb. Wiss.* 70, 151–160. doi: 10.1007/s41449-016-0024-9
- R Core Team (2025). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/> (Accessed July 20, 2025).
- Razin, Y. S., and Feigh, K. M. (2024). Converging measures and an emergent model: a meta-analysis of human-machine trust questionnaires. *ACM Trans. Hum.-Robot Inter.* 13, 1–41. doi: 10.1145/3677614
- Rossee, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Schaefer, K. (2013). *The Perception and Measurement of Human-Robot Trust* (Electronic Theses and Dissertations). Available online at: <https://stars.library.ucf.edu/etd/2688> (Accessed July 20, 2025).
- Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., and Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. *Hum. Fact.* 58, 377–400. doi: 10.1177/0018720816634228
- Schäfer, A., Esterbauer, R., and Kubicek, B. (2024). Trusting robots: a relational trust definition based on human intentionality. *Humanit. Soc. Sci. Commun.* 11:1412. doi: 10.1057/s41599-024-03897-3
- Schoeller, F., Miller, M., Salomon, R., and Friston, K. J. (2021). Trust as extended control: human-machine interactions as active inference. *Front. Syst. Neurosci.* 15:669810. doi: 10.3389/fnsys.2021.669810
- Speekenbrink, M. (2023). *An R Companion to Statistics: Data Analysis and Modelling.* Available online at: <https://mspeekenbrink.github.io/sdam-r-companion/> (Accessed June 30, 2025).
- Svebak, S., and Murgatroyd, S. (1985). Metamotivational dominance: a multimethod validation of reversal theory constructs. *J. Pers. Soc. Psychol.* 48, 107–116. doi: 10.1037/0022-3514.48.1.107
- van der Waa, J., Verdult, S., van den Bosch, K., van Diggelen, J., Haije, T., van der Stigchel, B., et al. (2021). Moral decision making in human-agent teams: human control and the role of explanations. *Front. Robot. AI* 8:640647. doi: 10.3389/frobt.2021.640647
- Wagner, A. R., Borenstein, J., and Howard, A. (2018). Overtrust in the robotic age. *Commun. ACM* 61, 22–24. doi: 10.1145/3241365
- Wang, L., He, X., and Chen, Y. (2016). Quantitative relationship model between workload and time pressure under different flight operation tasks. *Int. J. Ind. Ergon.* 54, 93–102. doi: 10.1016/j.ergon.2016.05.008

Werth, L., and Förster, J. (2015). *Regulatorischer Fokus. Zeitschrift für Sozialpsychologie*. Available online at: <https://econtent.hogrefe.com/doi/10.1024/0044-3514.38.1.33> (Accessed June 30 2025).

Wickens, C. D., Clegg, B. A., Vieane, A. Z., and Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Hum. Fact.* 57, 728–739. doi: 10.1177/0018720815581940

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Available online at: <https://ggplot2.tidyverse.org> (Accessed July 20, 2025).

Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *J. Am. Stat. Assoc.* 111, 1548–1563. doi: 10.1080/01621459.2016.1180986