



UvA-DARE (Digital Academic Repository)

My colleague is an AI! Trust differences between AI and human teammates

Georganta, E.; Ulfert, A.-S.

DOI

[10.1108/TPM-07-2023-0053](https://doi.org/10.1108/TPM-07-2023-0053)

Publication date

2024

Document Version

Final published version

Published in

Team Performance Management

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Georganta, E., & Ulfert, A.-S. (2024). My colleague is an AI! Trust differences between AI and human teammates. *Team Performance Management*, 30(1/2), 23-37. <https://doi.org/10.1108/TPM-07-2023-0053>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

My colleague is an AI! Trust differences between AI and human teammates

Trust
differences

23

Eleni Georganta

Research Group of Work and Organizational Psychology, Faculty of Social and Behavioural Sciences, University of Amsterdam, Amsterdam, The Netherlands, and

Anna-Sophie Ulfert

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

Received 7 July 2023
Revised 7 February 2024
12 February 2024
Accepted 15 February 2024

Abstract

Purpose – The purpose of this study was to investigate trust within human-AI teams. Trust is an essential mechanism for team success and effective human-AI collaboration.

Design/methodology/approach – In an online experiment, the authors investigated whether trust perceptions and behaviours are different when introducing a new AI teammate than when introducing a new human teammate. A between-subjects design was used. A total of 127 subjects were presented with a hypothetical team scenario and randomly assigned to one of two conditions: new AI or new human teammate.

Findings – As expected, perceived trustworthiness of the new team member and affective interpersonal trust were lower for an AI teammate than for a human teammate. No differences were found in cognitive interpersonal trust and trust behaviours. The findings suggest that humans can rationally trust an AI teammate when its competence and reliability are presumed, but the emotional aspect seems to be more difficult to develop.

Originality/value – This study contributes to human-AI teamwork research by connecting trust research in human-only teams with trust insights in human-AI collaborations through an integration of the existing literature on teamwork and on trust in intelligent technologies with the first empirical findings on trust towards AI teammates.

Keywords AI teammate, Trustworthiness, Cognitive interpersonal trust, Affective interpersonal trust, Trust behaviours

Paper type Research paper

© Eleni Georganta and Anna-Sophie Ulfert. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This research was supported by a funding from the laboratory for experimental research in economics of the Technical University of Munich (experimenTUM) and by the Society for Industrial and Organizational Psychology Foundation as part of the Visionary Circle.

Author note: Eleni Georganta and Anna-Sophie Ulfert have contributed equally to this work and share first authorship.



Introduction

Change is an ever-present reality for work teams and modern organisations in general, and teams therefore need to continuously adapt to varying conditions, such as membership changes or the introduction of new autonomous technologies (Maynard *et al.*, 2015). Given the rapid technological developments in the field of artificial intelligence (AI), adaptation demands are becoming increasingly challenging for teams. In some cases, autonomous technologies are becoming teammates (Larson and DeChurch, 2020; Seeber *et al.*, 2020), requiring teams and individual members to adapt to having AI colleagues (Ulfert-Blank *et al.*, 2023) and to continue operating effectively.

AI teammates are autonomous technologies that operate alongside humans, participate in cognitive decision-making and fulfil a distinct role that contributes to team performance. Human–AI teams (HAITs) are therefore a collection of human individuals and one or more AI agents that interact virtually, perform interdependent tasks and are embedded in an organisational system (Kozłowski and Ilgen, 2006; Ulfert *et al.*, 2023). The effectiveness of HAITs mainly depends on the level of trust among the human and AI teammates, given that trust reflects a fundamental factor for both team success and human–AI collaborations (Salas *et al.*, 2005; Sanders *et al.*, 2011). While the importance of trust has been recognised, the development of trust relationships among even human teammates remains challenging (Frazier *et al.*, 2013; McAllister, 1995), and when a new teammate is introduced – especially an AI teammate – existing team members need to develop trust towards the new addition at the same time, potentially, as trust towards a new autonomous technology. Team research has shown that a lack of prior history among teammates can lead to low trust and thus to lower transparency and acceptance of others (Grossman and Feitosa, 2018; Zand, 1972). Similarly, work on AI has found that lack of prior experience with autonomous agents can lead to low trust in the technology (Schaefer *et al.*, 2016; Ulfert and Georganta, 2020).

By integrating the trust and team literature into the context of human–AI collaborations, the goal of the present study is to provide a better understanding of trust in HAITs. To do this, we investigate whether trust is different when introducing a new AI teammate than a new human teammate. We expect perceived trustworthiness of the new teammate, cognitive and affective interpersonal trust towards the teammate and trust behaviours to be lower when introducing an AI than a human. Hence, our focus is not on the relationships between interpersonal trust, its antecedents (trustworthiness) and its outcomes (trust behaviours) in HAITs but rather on whether these aspects differ between HAITs and human-only teams.

With this research, we hope to contribute to the trust, teams and human–AI interaction literature in three ways. First, we apply trust theories to explore human–AI trust relationships and discuss the theoretical implications for the cognitive and affective dimensions of interpersonal trust. Second, we combine research on trust in teams with research on trust in technology and provide insights into trustworthiness and interpersonal trust within a new type of team. Third, we bring team research and human–AI interaction research closer together and present initial evidence of perceptions and intentions regarding an AI agent, both as a new teammate and as a new autonomous technology. We hope that our study will serve as a starting point for exploring the complex mechanisms of trust in HAITs, as recently called for by Ulfert *et al.* (2023), and for finding ways to design and introduce AI agents as part of a team.

Theoretical background

Over the past few decades, team and human–computer interaction scholars have increasingly argued that trust reflects a vital mechanism for the effective functioning of both human and human–technology relationships (Grossman and Feitosa, 2018; Sheridan, 2019).

Specifically, high levels of trust have been found to lead to high cohesion, satisfaction and learning within teams (Breuer *et al.*, 2016), and trust has been shown to have a positive impact on team performance, especially when interdependency among team members is high (De Jong *et al.*, 2016). Research on human–robot interactions has similarly found that trust in autonomous technologies is directly related to team effectiveness and performance (Lee and See, 2004).

In both human-only teams and HAITs, trust towards each individual teammate, which is referred to as interpersonal trust (McAllister, 1995), is a highly complex phenomenon. Here, we define interpersonal trust towards a human or AI teammate as “the willingness of a party [1] to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer *et al.*, 1995, p. 712). Furthermore, in line with prior work, we argue that interpersonal trust towards both human and AI teammates is multidimensional and integrates cognitive and affective components (Cook and Wall, 1980; Glikson and Woolley, 2020; McAllister, 1995; Webber, 2008). Finally, we propose that interpersonal trust towards a teammate is related to the perceptions of the trustworthiness of that teammate (Mayer *et al.*, 1995) and to trust behaviours towards them (Schoorman *et al.*, 2007).

High trust is a key factor for effective collaboration, and conversely, low trust can lead to detrimental outcomes, such as poor decision-making, limited information exchange, misunderstandings and personal conflicts (Häkkinen, 2004; Hartman, 2002). This is because – especially when forming a new team or when a new teammate is introduced – building trust can be difficult (Grossman and Feitosa, 2018). To build trust relationships, team members import trust-related information from previous similar situations, incorporate generalised expectations and collect information from the new environment (Wildman *et al.*, 2012). We argue that when an AI teammate is newly introduced, information from previous situations is limited and both expectations and similarities among teammates are relatively low. We, therefore, expect trustworthiness of a new teammate, cognitive and affective interpersonal trust towards that teammate and trust behaviours to be lower for a new AI teammate than a new human teammate (see Figure 1).

Trustworthiness perceptions of new teammates

Trust is a psychological state of a trustor (a person who can trust or distrust, e.g. an existing team member) towards a trustee (a person who can be trusted or distrusted, e.g. a new AI or human teammate) that entails the willingness to accept vulnerability based on positive expectations of trustworthiness (De Jong and Elfring, 2010). Trustworthiness is a multifaceted construct comprising three dimensions – ability, integrity and benevolence (Mayer *et al.*, 1995) – and thus when a new teammate is introduced, their ability, integrity and benevolence will impact whether they are perceived as trustworthy.

In a team, the dimension of ability can be divided into task-related (e.g. competence) and team-related (e.g. proactive behaviour; Breuer *et al.*, 2020) abilities. Specifically, ability

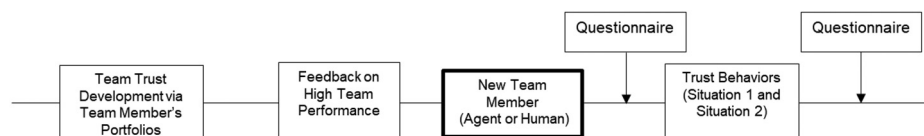


Figure 1.
Procedure of the study

Source: Figure by authors

captures the knowledge and expertise that a teammate needs to complete their tasks and the interpersonal and soft skills required for effective collaboration (Colquitt *et al.*, 2007). Integrity – the second dimension of trustworthiness – reflects the teammate’s credibility, sense of justice, moral standards and consistency (Fulmer and Gelfand, 2012). The third dimension – benevolence – comprises courtesy and both concern for and a positive orientation towards the team (Mayer *et al.*, 1995; Zolin *et al.*, 2004).

The ability and integrity of a new teammate can be determined by rational reasoning, but to assess benevolence, some interaction and emotional attachment are required (Mayer *et al.*, 1995). Ability relies mainly on facts (e.g. track records and information about history and performance), and integrity can be evaluated with the help of moral standards and a personal sense of fairness (Colquitt *et al.*, 2007; Jarvenpaa *et al.*, 1998), but a new teammate’s benevolence may be more difficult to assess immediately (Mayer *et al.*, 1995). Thus, in newly formed team relationships, ability, integrity and benevolence can have a substantial impact on trust (Aubert and Kelsey, 2003; Jarvenpaa *et al.*, 1998).

Given that the trustworthiness of a new teammate is also dependent on the context (Mayer *et al.*, 1995), we argue that the perceived ability, integrity and benevolence – and hence the perceived trustworthiness – of a new teammate will be different when that teammate is an AI than when it is a human. Specifically, we argue that an AI will be perceived as less trustworthy because the existing team members will be more uncertain in their judgements of an AI teammate’s competence and more confident in evaluating a human teammate’s ability and integrity (Glikson and Woolley, 2020). Moreover, we assume that existing team members will have less experience with AI than with human teammates and may therefore experience difficulties evaluating the AI teammate’s team orientation and benevolence. A lack of prior experience of interacting with a technological entity, such as AI agent, can produce high levels of uncertainty and thus misjudgements of an AI teammate’s ability, integrity and benevolence (Sanders *et al.*, 2011). Indeed, trust towards AI agents can be analogous, but not identical, to trust in humans (Culley and Madhavan, 2013; De Visser *et al.*, 2017) and may develop based on similar perceptions and interactions, but it can also differ due to cognitive bias, scepticism and irrational factors (Glikson and Woolley, 2020). Especially during the early phases of interaction, the perception and evaluation of an AI teammate’s trustworthiness may thus be impacted by factors unrelated to the AI’s actual trustworthiness, and we therefore hypothesise the following:

- H1. The perceived trustworthiness of a new team member will be lower for a new AI teammate than a new human teammate.

Interpersonal trust towards new teammates

According to McAllister (1995), interpersonal trust towards a teammate consists of two dimensions:

- (1) cognitive interpersonal trust, which is an individual’s or a team’s “confidence in the ability of others, yielding ascriptions of capability and reliability”; and
- (2) affective interpersonal trust, which is an individual’s or a team’s “faith in the trustworthy intentions of others” (p. 40).

When a new teammate is introduced, cognitive interpersonal trust is built on the knowledge available about the teammate’s competences, reliability and dependability (Costa *et al.*, 2018;

Luhmann, 1988; Schaubroeck *et al.*, 2011), and evidence suggests that, even in newly formed trust relationships, cognitive interpersonal trust can be high (Webber, 2008). However, affective interpersonal trust is built on emotional ties, care and concern between the existing team members and the new teammate (Al-Ani and Redmiles, 2009) and on the belief that these affective tendencies are reciprocated (Costa *et al.*, 2018; Lewis and Weigert, 1985). In contrast to cognitive interpersonal trust, affective interpersonal trust develops through prolonged interaction and may follow after the cognitive component (Webber, 2008). Research supports the two-dimensional nature of interpersonal trust, showing that both dimensions have a positive impact on team collaboration (Barczak *et al.*, 2010) and team performance (Erdem and Ozen, 2003).

We expect cognitive and affective interpersonal trust to be lower when a new AI teammate is introduced than a new human teammate. This is because, when introducing a new teammate, the existing team members will base their cognitive and affective interpersonal trust on previous team experiences (Grossman and Feitosa, 2018) and on their degree of similarity with the new teammate (see also social categorisation theory; Turner *et al.*, 1987). Lack of previous interaction with autonomous technologies and thus a lack of knowledge about an AI's capabilities and performance may lead to low interpersonal trust towards a new AI teammate (Sanders *et al.*, 2011; Turner *et al.*, 1987); being less similar to an AI teammate may also lead to expectations of fewer positive attributes and characteristics (Turner *et al.*, 1987). Consequently, evaluations of an AI teammate's cognitive and affective cues, intentions and goals may not be as accurate as evaluations of a human teammate (Grossman and Feitosa, 2018). Perhaps inevitably, developing feelings of similarity with AI teammates and building trusting relationships takes time (Ulfert and Georganta, 2020), and we therefore hypothesise the following:

- H2a. Cognitive interpersonal trust will be lower when a new AI teammate is introduced than a new human teammate.
- H2b. Affective interpersonal trust will be lower when a new AI teammate is introduced than a new human teammate.

Trust behaviours towards new teammates

Trustworthiness and cognitive and affective interpersonal trust refer to the internal states of the existing team members (trustors) towards a new teammate (trustee), whereas trust behaviours refer to “the observable interaction of the trustor with the trustee, where risk is taken by the trustor's dependence on the trustee in a certain situation, following upon a positive trust decision” (Rusman *et al.*, 2010, p. 837). Thus, the evaluation of a new teammate's trustworthiness and the emergent cognitive and affective interpersonal trust towards them are also related to the trust behaviours of the existing team members, and low degrees of similarity, less developed relationships and limited experience with an AI teammate will influence these behaviours (Kramer *et al.*, 2001). Because people are less familiar with being introduced to an AI teammate than a human teammate, trust behaviours are more difficult to generate (Cramton and Webber, 2005; Rusman *et al.*, 2010), and we therefore hypothesise the following:

- H3. When a new AI teammate is introduced, existing team members will demonstrate fewer trust behaviours than when a new human teammate is introduced.

Methods

Sample

To determine the sample size, we ran an *a priori* power analysis using G*Power (version 3.1.9.2; Faul *et al.*, 2007) with a power level of 0.95, an alpha-error level of 0.05 and an assumed medium effect size between the variables (Feitosa *et al.*, 2020). A sample size of 36 individuals for each of the two conditions (Condition A with an AI teammate, Condition B with a human teammate) was calculated. The final sample was therefore considered sufficient, with 127 subjects [54.3% female; mean (SD) age = 24.39 (4.13) years] – 59 in Condition A and 68 in Condition B. Of these, 74% were students, 7.1% employed full time, 15% employed part time and 5% unemployed. The sample was composed of 28 nationalities, with the majority German (57.4%), followed by Turkish (6.3%) and Russian (6.3%).

Design and experimental scenario

Using a between-subjects design, we conducted an online experiment in which the participants were introduced to a new randomly assigned teammate, either AI (Condition A) or human (Condition B). Using the open-source software oTree (Chen *et al.*, 2016), we developed an online team scenario; the participants were told that they were members of a five-member interdisciplinary team whose goal was to organise a new author's book release and subsequent book tour. The participants did not actually work in a real team nor execute tasks during the experiment – all the information received was part of a hypothetical scenario.

After the participants had received initial information about their team, the team members and the team's good performance thus far, they were told that one team member was to be replaced by either an AI (Condition A) or a human (Condition B) teammate. After the new teammate was introduced, we examined the differences between the two conditions regarding the perceived trustworthiness of the new teammate, cognitive and affective trust towards them and trust behaviours.

Procedure

The participants were told to assume that they were a marketing expert in a publishing company and about to start working in an interdisciplinary team with four other people: an editor, a designer, a data analyst and a financial adviser. They were also told that the team members were working together for the first time. Nevertheless, the participants did not actually interact with the team and did not complete any team tasks. To test the hypotheses, we assessed only the participants' perceptions based on the hypothetical information provided as part of the scenario.

The participants were asked to provide information about their hometown, favourite hobbies, how they approach problems and what they value most when working in a team. They were then asked to carefully read the online portfolios of their fellow team members to prepare for the project's kick-off meeting. The online portfolios described the four team members as having sufficient skills and abilities to make them experts in their respective roles. Each team member's portfolio also included answers to the same questions that the participants had been asked, creating intragroup similarities, and the information provided in the online portfolios supported the development of clear roles and team cohesion [2]. Next, the participants were informed that the project – after some weeks – was running well and that team performance was good.

The participants were then told that the data analyst had been unexpectedly transferred to another project and was to be replaced by a new data analyst, either an AI (Condition A)

or a human (Condition B). In both conditions, the participants were asked to read the online portfolio of the new teammate, which had an identical structure to the previous portfolios. In the online portfolio, the new teammate was referred to either as a “data analyst AI agent” (Condition A) or a “data analyst” (Condition B). At the same time, these portfolios did not include information related to the participants hometown and hobbies. There were no other differences between the conditions (see [Appendix](#) for online portfolios).

After reading the online portfolio of the new teammate, all participants were presented with two scenarios. In the first, they were told that the team needed to distribute its tasks to plan the book tour successfully, and in the second, the team had to unexpectedly change the original tour to meet the author’s new requirements. In both scenarios, the participants had to decide how the team should deal with the situations by selecting one of three possible options.

Before and after the scenarios were presented, the participants completed separate online questionnaires. Both assessed the perceived trustworthiness of the new teammate as well as cognitive and affective interpersonal trust; the second also assessed demographics (age, gender, nationality and employment status). Trust behaviours were measured by the participant’s selected option in the two presented scenarios, with each option reflecting a different level of trust behaviour (low, medium or high). The study procedure is presented in [Figure 1](#).

Measures

If not stated otherwise, all responses were given using a five-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree).

Perceived trustworthiness. We assessed the perceived trustworthiness of the new teammate (AI or human) with eight items (e.g. “Overall, my new team member is trustworthy”) from [Jarvenpaa et al. \(1998\)](#). Owing to low scale reliability, we removed two reverse-scored items (“There is no “team spirit” in my group with the new team member”, “There is a noticeable lack of confidence among those with whom I work”), which resulted in more reliable data.

Cognitive interpersonal trust. We assessed cognitive interpersonal trust with three items (e.g. “Given my team members’ track records, I see no reason to doubt their competence and preparation for the project’) from [McAllister \(1995\)](#).

Affective interpersonal trust. We assessed affective interpersonal trust with three items (e.g. “I can talk freely to my team members about difficulties I am having at the project and know that they will want to listen”) from [McAllister \(1995\)](#).

Trust behaviours. We assessed trust behaviours by asking the participants to decide how the team should act in two scenarios:

- (1) distributing the team tasks for planning the author’s book tour; and
- (2) reacting to an unexpected change of the original tour plan.

For each scenario, the participants were asked to select one of three possible options, with each reflecting either a low (e.g. supervisor distributes the tasks), medium (e.g. each member selects its own tasks) or high (e.g. collective decision about task distribution) level of trust behaviours. The scores were assessed on a scale from 0 (selecting the option that reflected a low level of trust behaviours in both scenarios) to 4 (selecting the option that reflected a high level of trust behaviours in both scenarios).

Data analysis

We used *t*-tests to test for differences between Condition A (AI teammate) and Condition B (human teammate) in the perceived trustworthiness of the new teammate (*H1*) and in cognitive (*H2a*) and affective (*H2b*) interpersonal trust. We used χ^2 tests to compare the frequencies with which the different options reflecting different levels of trust behaviours between the two conditions were selected (*H3*). SPSS 26 (IBM) was used to perform the *t*-tests and χ^2 tests. Means, standard deviations and correlations between the study variables are presented in [Table 1](#).

Results

As hypothesised, perceived trustworthiness of an AI teammate ($M = 3.75, SD = 0.60$) was significantly lower than that of a human teammate [$M = 4.02, SD = 0.60; t(125) = -2.58, p = 0.011$]. *H1* was therefore supported.

However, there were no significant differences in cognitive interpersonal trust between the introduction of an AI ($M = 4.24, SD = 0.69$) and the introduction of a human teammate [$M = 4.28, SD = 0.65; t(125) = -0.29, p = 0.767$]. *H2a* was therefore not supported.

As hypothesised, affective interpersonal trust was significantly lower when introducing a new AI teammate ($M = 3.58, SD = 0.93$) than a new human teammate [$M = 4.02, SD = 0.70; t(125) = -3.04, p < 0.01$]. *H2b* was therefore supported.

Finally, in contrast to expectations, no differences were found in the number of options selected that reflected low [$\chi^2(1) = 0.26, p = 0.862$], medium [$\chi^2(1) = 0.22, p = 0.635$] or high [$\chi^2(1) = 0, p = 1.000$] levels of trust behaviours between the AI teammate and human teammate conditions. *H3* was therefore not supported.

Discussion

Modern work environments are constantly changing, and both team ([Larson and DeChurch, 2020](#)) and human–AI interaction researchers ([Seeber et al., 2020](#)) have argued that AI agents will soon become our teammates. Given that trust will be required to adapt successfully to HAITs ([Salas et al., 2005; Ulfert and Georganta, 2020](#)), the goal of the present study was to explore trust when an AI teammate is introduced to an established team. We argued that the existing team members would perceive a new AI teammate as less trustworthy than a new human teammate, mainly due to a lack of experience in interacting with such technological entities. We further argued that interpersonal trust is multidimensional ([Mayer et al., 1995](#)), and that both cognitive and affective interpersonal trust would be lower for a new AI teammate due to a lack of similar team experiences and a low similarity with the AI. Finally, we argued that existing team members would display fewer trust behaviours when an AI teammate is introduced.

Table 1.
Cronbach's alpha, means, standard deviations and inter-correlations of the dependent variables

	α	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. Condition	–	–	–	–				
2. Trustworthiness	0.83	3.90	0.62	0.22*	–			
3. Cognitive trust	0.82–0.89	4.27	0.67	0.03	0.47**	–		
4. Affective trust	0.74–0.84	3.82	0.84	0.26**	0.59**	0.50**	–	
5. Trust behaviors	–	0.58	0.73	–0.08	–0.12	–0.07	–0.10	–

Notes: * $p < 0.05$; ** $p < 0.001$

Source: Table by authors

We conducted an online experiment to test our assumptions, and as expected, perceived trustworthiness and affective interpersonal trust were lower for an AI teammate than a human teammate. However, no differences were found in cognitive interpersonal trust or trust behaviours. Our study contributes to team and human–AI interaction research by extending theories of trust within human-only teams to trust development in HAITs, integrating the literature on teamwork with the literature on trust in technology and providing the first empirical findings on trust towards AI teammates, as recently called for by [Ulfert et al. \(2023\)](#).

We found that when a new teammate is introduced, existing team members evaluated its trustworthiness as lower when it is an AI agent than when it is a human. A possible explanation is that the existing team members were simply more cautious or insecure in making the evaluation. Perceived trustworthiness includes not only an AI teammate's ability, integrity and benevolence but also expectations regarding its behaviour ([Fulmer and Gelfand, 2012](#)), and previous research suggests that having no prior experience interacting with autonomous technologies, such as AI agents, can result in uncertainty about both competence ([Glikson and Woolley, 2020](#)) and behaviour.

Contradicting our reasoning, we found that cognitive interpersonal trust of an AI and a human teammate were not significantly different. It is possible that the assessment of cognitive cues was based solely on the information available, which was identical for the AI and human teammates, yet lack of experience in interacting with autonomous agents did not seem to influence the evaluation of cognitive interpersonal trust, resulting in an accurate rating of the AI agent's competences. This is in line with prior work that found that cognitive interpersonal trust derives from rational reasoning and thus is evaluated with a sense of fairness ([Colquitt et al., 2007](#)).

Unlike cognitive interpersonal trust, affective interpersonal trust was, as hypothesised, lower after introducing an AI teammate than a human one, which reflects previous findings that interpersonal trust can be differentiated into cognitive and affective components ([McAllister, 1995](#)). The low degree of similarity between the team members and the AI and the consequent categorisation of the AI as part of an outgroup ([Turner, 2010](#); [Turner et al., 1987](#)) may have negatively impacted affective interpersonal trust, which may have impacted expectations about the AI agent's ability to show affective behaviours, such as care. Perhaps critically, a perception of the AI teammate as primarily a technological entity and not “humanlike” may have intensified the sense of dissimilarity between the team members and the AI; as [Przegalinska et al. \(2019\)](#) noted, “A crucial part of trust is related to anthropomorphization [...] the process of anthropomorphization is not only about the attribution of superficial human characteristics but most importantly this essential one: a humanlike mind” (p. 789).

Our results also unexpectedly showed no differences in trust behaviours. It is possible that a rational evaluation of the AI teammate's competences shaped subsequent trust behaviours more than any affective component. Prior research has shown that, in early trust relationships, rational trust elements can precede affective ones and thus can have a stronger impact on team outcomes, such as performance ([Chua et al., 2008](#); [Hempel et al., 2009](#)). Another possible explanation is that the team lifecycle that was described was too short to impact the trust behaviours of existing team members; as [Rusman et al. \(2010\)](#) suggested, in new situations, the absence of prior history and bonding results in trust that is not “thick”, thus allowing for immediate trust behaviours towards a newcomer.

Limitations

The current study has several limitations that should be considered when interpreting the findings and planning future research. First, the scenarios did not allow for any interaction

and were, in that respect, not comparable to a real team situation. Nevertheless, they do offer a first indication of how individuals perceive AI agents when introduced as teammates. Second, the experiment investigated only the perceptions and behavioural intentions of a single team member and did not explore trust dynamics as a result of interactions among multiple team members, one of whom is a new teammate (e.g. chatting). For instance, computer-mediated communication among team members would have made the situation more realistic and could have impacted the development of trust to a significant extent (Tucker *et al.*, 2023). Third, the short duration of the experiment limited the team's perceived life cycle and may not have allowed for a sense of affective interpersonal trust to develop. Fourth, although we assessed trust behaviours using the decisions the participants made, these reflected behavioural intentions rather than necessarily the consequences of trust. We therefore suggest that future research should investigate real HAITs over a longer period and assess trust behaviours directly.

Implications for theory, research and practice

Our research is a starting point for further explorations of trust in HAITs and for identifying relevant factors and mechanisms that contribute to the development of trust towards AI teammates. Specifically, our evidence suggests that human-only team research can inform our understanding of HAITs (O'Neill *et al.*, 2023), but further investigation is required to determine whether existing trust theories need expansion (Ulfert *et al.*, 2023). Specifically, there is a need to gain a better understanding of the distinction between cognitive and affective interpersonal trust and to clarify whether and to what extent the two trust dimensions are needed for HAIT success. For example, affective computing could make an AI agent more capable of understanding human emotions and thus of reacting more appropriately, but we do not yet know how affective interpersonal trust impacts human–AI teamwork and whether implementing such features would improve teamwork more than working with a rational AI (Seeber *et al.*, 2020).

Furthermore, there is a need to gain a deeper understanding of how interpersonal trust is related to perceived trustworthiness, subsequent trust behaviours and decision-making in HAITs. Although our work adds to the limited empirical work on HAITs from an organisational psychology perspective (O'Neill *et al.*, 2023), the study focused only on whether differences in trust perceptions and trust behavioural intentions between human-only teams and HAITs exist. Furthermore, it remains unclear whether additional factors, such as the team setting, may impact these differences, the relationships between the constructs and their impact on team trust (Morrisette and Kisamore, 2020). Would teams react the same under high pressure (e.g. surgical operating theatre) and low pressure (e.g. daily team meeting)? We therefore hope that our findings can encourage and inspire future research to provide better insights into trust in HAITs.

Our findings may also offer some practical guidance for designing and introducing AI teammates into the workplace, paving the way for more effective and harmonious human–AI collaborations. To promote the perceived trustworthiness of AI teammates, we recommend the integration of Explainable AI (Gade *et al.*, 2020), which can provide insights into AIs' decision-making processes and thus demonstrate both their capabilities (rational component) and their care for the team (relational component). Furthermore, we recommend training people to work with AI teammates to shape their perceptions of AIs as teammates and increase their familiarity and positive experiences, which has been shown to foster trust and collaboration within HAITs (Johnson *et al.*, 2023).

Conclusion

AI agents may soon start becoming our teammates, and trust will be required for effective human–AI collaborations. Our findings suggest that perceived trustworthiness and affective interpersonal trust towards a new AI agent are more difficult to develop than towards a new human teammate. However, cognitive interpersonal trust – a more rational evaluation – and trust behaviours do not seem to differ between an AI teammate and a human teammate. Further research is needed to investigate whether both cognitive and affective interpersonal trust are required to fully trust an AI teammate – both as a team member and as a technology – and how these trust dimensions are impacted by perceived trustworthiness and in turn impact trust behaviours after an HAIT has worked together for some time.

Notes

1. As our focus is on interpersonal trust, by “party”, we are referring here to a single teammate.
2. A pilot study (N = 61 individuals) showed that role clarity and intragroup similarities were rated as significantly higher after reading the online portfolios.

References

- Al-Ani, B. and Redmiles, D. (2009), “Trust in distributed teams: support through continuous coordination”, *IEEE Software*, Vol. 26 No. 6, pp. 35-40.
- Aubert, B.A. and Kelsey, B.L. (2003), “Further understanding of trust and performance in virtual teams”, *Small Group Research*, Vol. 34 No. 5, pp. 575-618.
- Barczak, G., Lassk, F. and Mulki, J. (2010), “Antecedents of team creativity: an examination of team emotional intelligence, team trust and collaborative culture”, *Creativity and Innovation Management*, Vol. 19 No. 4, pp. 332-345.
- Breuer, C., Hüffmeier, J. and Hertel, G. (2016), “Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators”, *Journal of Applied Psychology*, Vol. 101 No. 8, p. 1151.
- Breuer, C., Hüffmeier, J., Hibben, F. and Hertel, G. (2020), “Trust in teams: a taxonomy of perceived trustworthiness factors and risk-taking behaviors in face-to-face and virtual teams”, *Human Relations*, Vol. 73 No. 1, pp. 3-34.
- Chen, D.L., Schonger, M. and Wickens, C. (2016), “oTree—an open-source platform for laboratory, online, and field experiments”, *Journal of Behavioral and Experimental Finance*, Vol. 9, pp. 88-97.
- Chua, R.Y.J., Ingram, P. and Morris, M.W. (2008), “From the head and the heart: Locating cognition-and affect-based trust in managers’ professional networks”, *Academy of Management Journal*, Vol. 51 No. 3, pp. 436-452.
- Colquitt, J.A., Scott, B.A. and LePine, J.A. (2007), “Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance”, *Journal of Applied Psychology*, Vol. 92 No. 4, p. 909.
- Cook, J. and Wall, T. (1980), “New work attitude measures of trust, organizational commitment and personal need non-fulfilment”, *Journal of Occupational Psychology*, Vol. 53 No. 1, pp. 39-52.
- Costa, A.C., Fulmer, C.A. and Anderson, N.R. (2018), “Trust in work teams: an integrative review, multilevel model, and future directions”, *Journal of Organizational Behavior*, Vol. 39 No. 2, pp. 169-184.
- Cramton, C.D. and Webber, S.S. (2005), “Relationships among geographic dispersion, team processes, and effectiveness in software development work teams”, *Journal of Business Research*, Vol. 58 No. 6, pp. 758-765.

- Culley, K.E. and Madhavan, P. (2013), "A note of caution regarding anthropomorphism in HCI agents", *Computers in Human Behavior*, Vol. 29 No. 3, pp. 577-579.
- De Jong, B.A. and Elfring, T. (2010), "How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort", *Academy of Management Journal*, Vol. 53 No. 3, pp. 535-549.
- De Jong, B.A., Dirks, K.T. and Gillespie, N. (2016), "Trust and team performance: a meta-analysis of main effects, moderators, and covariates", *Journal of Applied Psychology*, Vol. 101 No. 8, pp. 1134-1150.
- De Visser, E.J., Monfort, S.S., Goodyear, K., Lu, L., O'Hara, M., Lee, M.R., Parasuraman, R. and Krueger, F. (2017), "A little anthropomorphism goes a long way: effects of oxytocin on trust, compliance, and team performance with automated agents", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 59 No. 1, pp. 116-133.
- Erdem, F. and Ozen, J. (2003), "Cognitive and affective dimensions of trust in developing team performance", *Team Performance Management: An International Journal*, Vol. 9 Nos 5/6.
- Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007), "G* power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences", *Behavior Research Methods*, Vol. 39 No. 2, pp. 175-191.
- Feitosa, J., Grossman, R., Kramer, W.S. and Salas, E. (2020), "Measuring team trust: a critical and meta-analytical review", *Journal of Organizational Behavior*, Vol. 41 No. 5, pp. 479-501.
- Frazier, M.L., Johnson, P.D. and Fainshmidt, S. (2013), "Development and validation of a propensity to trust scale", *Journal of Trust Research*, Vol. 3 No. 2, pp. 76-97.
- Fulmer, C.A. and Gelfand, M.J. (2012), "At what level (and in whom) we trust: trust across multiple organizational levels", *Journal of Management*, Vol. 38 No. 4, pp. 1167-1230.
- Gade, K., Geyik, S., Kenthapadi, K., Mithal, V. and Taly, A. (2020), "Explainable AI in industry: Practical challenges and lessons learned", Companion proceedings of the Web Conference, pp. 303-304.
- Glikson, E. and Woolley, A.W. (2020), "Human trust in artificial intelligence: Review of empirical research", *Academy of Management Annals*, Vol. 14 No. 2.
- Grossman, R. and Feitosa, J. (2018), "Team trust over time: modeling reciprocal and contextual influences in action teams", *Human Resource Management Review*, Vol. 28 No. 4, pp. 395-410.
- Häkkinen, P. (2004), "What makes learning and understanding in virtual teams so difficult?", *CyberPsychology and Behavior*, Vol. 7 No. 2, pp. 201-206.
- Hartman, F.T. (2002), *The Role of Trust in Project Management. The Frontiers of Project Management Research*, Project Management Institute, Newtown Square, PA.
- Hempel, P.S., Zhang, Z. and Tjosvold, D. (2009), "Conflict management between and within teams for trusting relationships and performance in China", *Journal of Organizational Behavior*, Vol. 30 No. 1, pp. 41-65.
- Jarvenpaa, S.L., Knoll, K. and Leidner, D.E. (1998), "Is anybody out there? Antecedents of trust in global virtual teams", *Journal of Management Information Systems*, Vol. 14 No. 4, pp. 29-64.
- Johnson, C.J., Demir, M., McNeese, N.J., Gorman, J.C., Wolff, A.T. and Cooke, N.J. (2023), "The impact of training on human-autonomy team communications and trust calibration", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 65 No. 7, pp. 1554-1570.
- Kozlowski, S.W.J. and Ilgen, D.R. (2006), "Enhancing the effectiveness of work groups and teams", *Psychological Science in the Public Interest*, Vol. 7 No. 3, pp. 77-124.
- Kramer, R.M., Hanna, B.A., Su, S. and Wei, J. (2001), "Collective identity, collective trust, and social capital: linking group identification and group cooperation", *Groups at Work: Theory and Research*, Vol. 173, p. 196.

-
- Larson, L. and DeChurch, L. (2020), "Leading teams in the digital age: four perspectives on technology and what they mean for leading teams", *The Leadership Quarterly*, Vol. 31 No. 1, pp. 1-18.
- Lee, J.D. and See, K.A. (2004), "Trust in automation: designing for appropriate reliance", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 46 No. 1, pp. 50-80, doi: [10.1518/hfes.46.1.50_30392](https://doi.org/10.1518/hfes.46.1.50_30392).
- Lewis, J.D. and Weigert, A. (1985), "Trust as a social reality", *Social Forces*, Vol. 63 No. 4, pp. 967-985.
- Luhmann, N. (1988), "Familiarity, confidence, trust: problems and alternatives", in Gambetta, D.G. (Ed.), *Trust: Making and Breaking Cooperative Relations*, Basil Blackwell, New York, NY, pp. 94-107.
- McAllister, D.J. (1995), "Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations", *Academy of Management Journal*, Vol. 38 No. 1, pp. 24-59.
- Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995), "An integrative model of organizational trust", *The Academy of Management Review*, Vol. 20 No. 3, pp. 709-734.
- Maynard, M.T., Kennedy, D.M. and Sommer, S.A. (2015), "Team adaptation: a fifteen-year synthesis (1998–2013) and framework for how this literature needs to 'adapt' going forward", *European Journal of Work and Organizational Psychology*, Vol. 24 No. 5, pp. 1-26, doi: [10.1080/1359432X.2014.1001376](https://doi.org/10.1080/1359432X.2014.1001376).
- Morrisette, A.M. and Kismore, J.L. (2020), "Trust and performance in business teams: a meta-analysis", *Team Performance Management: An International Journal*, Vol. 26 Nos 5/6, pp. 287-300.
- O'Neill, T.A., Flathmann, C., McNeese, N.J. and Salas, E. (2023), "Human-autonomy teaming: need for a guiding team-based framework?", *Computers in Human Behavior*, Vol. 146, p. 107762.
- Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P. and Mazurek, G. (2019), "In bot we trust: a new methodology of chatbot performance measures", *Business Horizons*, Vol. 62 No. 6, pp. 785-797.
- Rusman, E., Van Bruggen, J., Sloep, P. and Koper, R. (2010), "Fostering trust in virtual project teams: towards a design framework grounded in a Trust worthiness Antecedents (TWAN) schema", *International Journal of Human-Computer Studies*, Vol. 68 No. 11, pp. 834-850.
- Salas, E., Sims, D.E. and Burke, C.S. (2005), "Is there a 'big five' in teamwork?", *Small Group Research*, Vol. 36 No. 5, pp. 555-599.
- Sanders, T., Oleson, K.E., Billings, D.R., Chen, J.Y.C. and Hancock, P.A. (2011), "A model of human-robot trust: theoretical model development", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55 No. 1, pp. 1432-1436, doi: [10.1177/1071181311551298](https://doi.org/10.1177/1071181311551298).
- 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", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 58 No. 3, pp. 377-400, doi: [10.1177/0018720816634228](https://doi.org/10.1177/0018720816634228).
- Schaubroeck, J., Lam, S.S.K. and Peng, A.C. (2011), "Cognition-based and affect-based trust as mediators of leader behavior influences on team performance", *Journal of Applied Psychology*, Vol. 96 No. 4, p. 863.
- Schoorman, F.D., Mayer, R.C. and Davis, J.H. (2007), *An Integrative Model of Organizational Trust: Past, Present, and Future*, Academy of Management Briarcliff Manor, New York, NY, p. 10510.
- Seeber, I., Waizenegger, L., Seidel, S., Morana, S., Benbasat, I. and Lowry, P.B. (2020), "Collaborating with technology-based autonomous agents", *Internet Research*, Vol. 30 No. 1.
- Sheridan, T.B. (2019), "Extending three existing models to analysis of trust in automation: Signal detection, statistical parameter estimation, and model-based control", *Human Factors*, Vol. 61 No. 7, p. 18720819829951.
- Tucker, C., Olsen, B. and Hale, R.T. (2023), "Trust and commitment: a comparative study of virtual team communication across industries", *Team Performance Management: An International Journal*, Vol. 29 Nos 1/2, pp. 152-165.

- Turner, J.C. (2010), "Social categorization and the self-concept: a social cognitive theory of group behavior".
- Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D. and Wetherell, M.S. (1987), *Rediscovering the Social Group: A Self-Categorization Theory*, Basil Blackwell.
- Ulfert-Blank, A.S., Georganta, E., Tielman, M. and Oron-Gilad, T. (2023), "Piecing together the puzzle: understanding trust in human-AI teams", *In CEUR Workshop Proceedings*, Vol. 3456, pp. 169-174.
- Ulfert, A.-S. and Georganta, E. (2020), "A model of team trust in Human-Agent teams", *ACM International Conference Proceeding Series*.
- Ulfert, A.S., Georganta, E., Centeio Jorge, C., Mehrotra, S. and Tielman, M. (2023), "Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework", *European Journal of Work and Organizational Psychology*, pp. 1-14.
- Webber, S.S. (2008), "Development of cognitive and affective trust in teams: a longitudinal study", *Small Group Research*, Vol. 39 No. 6, pp. 746-769.
- Wildman, J.L., Shuffer, M.L., Lazzara, E.H., Fiore, S.M., Burke, C.S., Salas, E. and Garven, S. (2012), "Trust development in swift starting action teams: a multilevel framework", *Group and Organization Management*, Vol. 37 No. 2, pp. 137-170.
- Zand, D.E. (1972), "Trust and managerial problem solving", *Administrative Science Quarterly*, Vol. 17 No. 2, pp. 229-239.
- Zolin, R., Hinds, P.J., Fruchter, R. and Levitt, R.E. (2004), "Interpersonal trust in cross-functional, geographically distributed work: a longitudinal study", *Information and Organization*, Vol. 14 No. 1, pp. 1-26.

Further reading

- IBM (2024), "IBM SPSS statistics, version 26.0", IBM Corp.
- Levin, D.Z., Whitener, E.M. and Cross, R. (2006), "Perceived trustworthiness of knowledge sources: the moderating impact of relationship length", *Journal of Applied Psychology*, Vol. 91 No. 5, p. 1163.

Appendix. Information about new teammate for Condition A (AI) and Condition B (human)

Your Data Analyst, Stefanie Brown, has been unexpectedly transferred to another project. To make sure that your team completes its tasks without lacking expertise, the Project Manager decided to *immediately replace Stefanie's position*. Fortunately, the company's board is prepared for such occurrences and presents you your *new AI teammate (Condition A)/new teammate (Condition B) Dawn, an AI Data Analyst (Condition A)/a Data Analyst (Condition B)* with specific experience in handling the company's data.

As you are interested to know more about Dawn, you decided to have a look at your AI teammate's (Condition A)/teammate's (Condition B) online portfolio. The portfolio includes information about Dawn's top three qualities, the special certifications or honours your teammate has received, relevant information about their work at the company, as well as feedback from previous colleagues.

Description AI teammate (Condition A)/teammate (Condition B) – AI Data Analyst (Condition A)/Data Analyst (Condition B) Dawn

Top three qualities of Dawn:

- (1) Dawn presents their data analysis in an understandable and well-structured way.
- (2) Dawn handles data in a safe and secure way.
- (3) Dawn effectively communicates relevant information to the right person.

Certifications: Dawn was the winner for the best data analyst at the *Big Insight Data and AI Innovation Awards* in 2023.

Relevant information about Dawn's work

- Dawn makes accurate predictions based on previous data. Dawn learns from feedback to present improved solutions.
- Dawn integrates all important information regarding different locations and potential audiences to plan the book tour in the most profitable way.

Dawn has received the following feedback

Dawn is a very efficient AI teammate (Condition A)/teammate (Condition B). Dawn operates transparently and is easily integrated in a new team. Dawn handles data systematically and delivers reliable outcomes.

Source: By authors

Corresponding author

Eleni Georganta can be contacted at: e.georganta@uva.nl

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com