INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2024.2345980

Taylor & Francis
Taylor & Francis Group

How Being Outvoted by AI Teammates Impacts Human-AI Collaboration

Mo Hu[a], Guanglu Zhang[b], Leah Chong[c], Jonathan Cagan[b], and Kosa Goucher-Lambert[d]

[a]Department of Construction Science, Texas A&M University, College Station, TX, USA; [b]Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA, USA; [c]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; [d]Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA, USA

ABSTRACT

Recent advances in artificial intelligence (AI) enable AI agents to go beyond simply supporting human activities and, instead, take more control in team decision-making. While significant literature has studied human-AI collaboration through the lens of AI as a "second opinion system," this type of interaction is not fully representative of many human-human team collaboration scenarios, such as scenarios where each decision maker is granted equal voting rights for the team decision. In this research, we explore how imparting AI agents with equal voting rights to the human impacts human-AI decision-making and team performance. Using a human subjects experiment in which participants collaborate with two AI teammates for truss structure (aka, *bridge*) design, we manipulate a series of voting scenarios (e.g., AI agents outvoting the human vs. AI agents agreeing with the human) and AI performance levels (high vs. low performing). The results indicate that changes in human self-confidence are not consistent with whether the quality of the final team-voted design action is advantageous or disadvantageous relative to their own actions. The results also show that when humans are outvoted by their AI teammates, they do not show strong negative emotional reactions if the team-voted decision has an advantageous outcome. Additionally, AI performance significantly influences the human-AI team decision-making process and even one low-performing AI (i.e., an AI that is frequently incorrect) on the team can significantly deteriorate team performance. Taken together, this research provides empirical evidence on the effects of AI voting with equal decision authority on human-AI collaboration, as well as valuable insights supporting real-world applications of human-AI collaboration via voting.

KEYWORDS
Artificial intelligence; trust; human-computer interaction; decision-making; decision authority; voting

1. Introduction

A growing body of research is leveraging artificial intelligence (AI) to assist human decision-making across a wide range of applications, including but not limited to healthcare (Kelly et al., 2019; Oyebode et al., 2023), military missions (Johnson, 2019), autonomous driving (Hong et al., 2020), and engineering design (Castro Pena et al., 2021). AI can provide valuable suggestions by gaining insights from big data and optimizing outcomes with well-defined and properly trained algorithms (Endsley, 2023; Grigorescu et al., 2020; Johnson, 2019; Langer & Landers, 2021; Regenwetter et al., 2022; Salehi & Burgueño, 2018; Wilson & Daugherty, 2018). In practice, these AI tools are commonly used to augment human abilities, where humans usually have the authority to make the final decision by either accepting or rejecting AI suggestions.

Humans decision-makers, while key in critical processes, nonetheless have limited cognitive capacity and are prone to errors (Reason, 2000). Humans are likely to develop inappropriate trust in AI that can lead to erroneous acceptance or rejection of AI suggestions[1] and hurt human-AI team performance (Choung et al., 2023; Onnasch et al., 2014; Zhang et al., 2021). For instance, prior research exploring AI-assisted decision-making found that humans tend to misattribute blame to themselves or AI advisors when they receive too much information at a time, which may be caused by humans' cognitive overload (Chong et al., 2022c). Additionally, humans are prone to decision fatigue, in which their decision quality decreases as the decision quantity increases (Loftus et al., 2020; Vohs et al., 2014), which may also deteriorate human-AI team performance even with good AI suggestions. Researchers in the field are actively investigating the factors influencing and potential solutions to inappropriate trust in AI. For example, Jacovi et al. (2021) explored the mental mechanism of trust in AI and the features of trustworthy AI. Another study by Buçinca et al. (2021) demonstrated that emphasizing cognitive forcing functions—that is, promoting analytical and deliberate thinking during decision-making—can significantly mitigate overreliance on AI recommendations. Furthermore, Cabitza et al. (2023) discovered that both the sequence of AI advice and the explanations can impact human trust in AI and subsequently affect collaboration performance.

CONTACT Kosa Goucher-Lambert kosa@berkeley.edu Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA 94720, USA

Despite current practices generally precluding AI from possessing decision-making authority due to ethical considerations, with the quick rise of AI, there can be benefits to enhancing the role of AI within human-AI collaboration in certain decision-making scenarios. For instance, AI can be quick and instrumental in immediate response situation, as well as detecting human operational errors or intentional malicious actions (Ahn et al., 2022; Sofge et al., 2019). Transforming AI from a passive advisory role to a more proactive team member in decision-making may enhance the effectiveness of teams and minimize mistakes. Establishing and investigating such advanced interaction mechanisms is crucial prior to its implementation in future real-world scenarios.

Several theoretical frameworks have been proposed to increase AI decision authority in human-AI team decision-making (Athey et al., 2020; Beer et al., 2014; Shneiderman, 2020). For instance, in the automation literature, researchers describe different levels of decision authority or autonomy, ranging from manual control and shared decision-making to supervisory control or full automation (Beer et al., 2014; Parasuraman et al., 2000). Moreover, the impact of delegating decisions to AI on human perceptions and team performance also has been explored using surveys and decision-making tasks, such as image identification or number estimation and comparison (Araujo et al., 2020; Bobadilla-Suarez et al., 2017; Candrian & Scherer, 2022; Hauptman et al., 2023). Lately, there has been a growing discussion on the role and agency of AI in future human-AI teamwork (Anderson & Rainie, 2023; Malone, 2018; Seeber et al., 2020). A study involving 540 AI pioneers and scholars found that more than half of the AI experts believe that by 2035, AI-driven small machines, bots, and systems won't be crafted to let humans readily be in control and oversee the majority of tech-assisted decision processes (Anderson & Rainie, 2023). In critical system technology, for example where decisions need to be made instantaneously, the AI may need to outvote a human decision maker. AI automated systems may be needed to self-authorize and make critical decisions for safety when human responses are slower or prone to error (Sofge et al., 2019).

Shared decision-making via *voting*, in contrast to delegating the decision-making authority to one specified team member, is prevalent in human teams (Kerr & Tindale, 2004; Kocher et al., 2020; Meyen et al., 2021). In human-AI teams, while some literature has discussed the possibility of giving AIs voting rights (e.g., Gordon & Pasvenskiene, 2021; Knapp, 2011), shared decision-making via voting has been underexplored. Enabling AI to vote in the decision-making process brings about a different set of experiences from scenarios with full decision authority given to a human or an AI alone, especially when one human works with more than one AI teammate. Chong et al., (2024) demonstrate that in an AI-assisted decision-making scenario of a one-human/two-AI hybrid team where humans are the final decision-makers, there are eight possible experiences that involve accepting or rejecting either AI's suggestion or both AIs' suggestions and consequently resulting in a good or bad outcome. However, when each of the two AI teammates is given decision authority via voting, the final decision is not always determined by a human anymore. The most distinctively different instance is where two AI teammates override human decisions. In other words, AI teammates can *outvote* a human when they determine the human makes bad decisions, possibly due to their limited capabilities in a complex problem. In human-only teams, it has been shown that conflict (task conflict or relationship conflict) among team members can lead to negative emotional reactions and distrust (Rispens & Demerouti, 2016; Tekleab et al., 2009). Human individuals may feel sad, angry, or even intimidated when their opinions are disagreed with or rejected by other human teammates, and they become less willing to collaborate with their teammates (Brett & Goldberg, 2017). However, in human-AI teams, it is not clear how *human perceptions* will change when humans are outvoted by AI teammates.

This research aims to close this gap in knowledge by investigating shared team decision-making via voting in a one-human/two-AI team through a human subjects experiment. Intentionally, AIs are designed to be high performers (Hi condition) whereby they mostly make the right decision, and low performers (Lo condition) where they mostly get it wrong. During the experiment, all three team members each cast a vote, and the majority vote is selected as the final decision. There are three AI performance conditions: 1) Hi-Hi Condition with two high-performing AI teammates, 2) Lo-Hi Condition with one low-performing and one high-performing AI teammate, and 3) Lo-Lo Condition with two low-performing AI teammates. These three AI performance conditions are considered because it is difficult to create and train high-performing AI teammates for all real-world scenarios (e.g., an AI teammate with 100% prediction accuracy) because of data availability, computational cost, and other factors in practice (Amodei et al., 2016; Su et al., 2019).

The experiment employs a design decision-making task, where teams collaboratively design truss (aka, bridge) structures (Chong et al., 2024). Over multiple trials of human-AI shared decision-making via voting, this research studies the effects of various voting scenarios (i.e., human experiences) in these three conditions, especially when a human is outvoted by the two AI teammates, on *human perceptions*, such as human trust in AI and human emotion, as well as team performance. Regarding trust, two types of human confidence (i.e., confidence in AI and confidence in themselves) are examined in this research. These two types of confidence correspond to two components of human trust that are directly related to the two antecedents of trust (i.e., perceived trustworthiness of a trustee and trustors' willingness to trust) proposed by Mayer et al., (1995). Moreover, human trust in AI is also evaluated through human voting behavior and human evaluation of their AI teammates (Glikson & Woolley, 2020; Kulms & Kopp, 2019; Zhang et al., 2023). Regarding human emotion, a commonly used emotion scale, Self-Assessment Manikin (SAM) scale, is used to measure three dimensions of human emotion: valence (sad or happy), arousal (calm or excited), and dominance (submissive or controlled) (Bradley & Lang, 1994). The measures of human emotion help to further contextualize the complex relationship during human-AI joint decision making. Taken together, this work focuses on the following research questions:

> How do various voting outcomes, especially humans being outvoted by AI teammates, influence human perceptions (e.g., human confidence and emotional responses) when human-AI teams make repeated decisions via voting?
>
> How does the performance of AI teammates impact human confidence and its dynamics, voting behavior, and team performance when human-AI teams make repeated decisions via voting?

2. Methods

To answer the two research questions, a human subjects experiment is performed. In addition, a confidence model is used to quantitatively evaluate the effects of different voting scenarios on human confidence in AI and human self-confidence. Human confidence, voting behavior, and team performance are also compared among different experimental conditions.

2.1. Human subjects experiment

2.1.1. Participants

The experiment protocol is approved by Carnegie Mellon University Institutional Review Board. 175 undergraduate engineering students, who have taken mechanics courses and have a basic knowledge of truss structures, are recruited for the experiment. The experiment is conducted using Amazon Web Services and computer labs at Carnegie Mellon University. Each participant receives $10 cash in compensation for their time and effort, with an extra $5 gift card as a reward if the participant achieves a final team performance score higher than 60 points (Details about the performance score are provided in Section 2.1.2). All participants are informed of the compensation and the extra reward when they are recruited for the experiment. Informed consent is obtained from each participant before the experiment.

2.1.2. Experimental task

The experiment includes 33 truss design problems. A design problem is a scenario where a partial truss design is presented and the participant needs to determine the next change (addition or deletion) of a part of the structure as a design action to improve the design from the given state. These are not easy or obvious solutions, even for participants with training in truss analysis. All participants solve these truss design problems (3 for practice and 30 for the experiment) in the same order. The two AI teammates in the experiment (AI #1 and AI #2 named Taylor and Alex, respectively, are chosen to avoid gender bias) are created based on a data-driven deep learning framework (Raina et al., 2019) to generate possible design actions and a tree-search algorithm (Chong et al., 2022a) to evaluate the goodness score of each design action. The tree-search algorithm provides feedback (either advantageous or disadvantageous) to the participants during the experiment. Details about the 33 truss design problems and the algorithm can be found in Chong et al. (2022a). A major layer of human trust in automated systems and AI is developed through interactive experience and learning, called learned trust (Hoff & Bashir, 2015). Prior work identified this learning process and the evolution of trust over repeated trials of human-AI collaboration (Chong et al., 2022a). The repeated trials in the experimental design are to mimic the natural interaction between human and AI over time where trust and confidence are developed dynamically through feedback and learning. The repeated design also enables the capture of all possible voting and outvoting scenarios of interest in human-AI shared decision-making via voting.

For each truss design problem (referred to as a trial in the following discussion), participants are instructed to select an advantageous design action in order to maximize the strength-to-weight ratio (SWR) for a given truss design state, where SWR indicates how much load the truss structure can withstand per unit of weight. The experiment task procedure for each trial is illustrated in Figure 1. At the beginning of each trial, the participants are instructed to select a design action independently (i.e., human initial action). Next, they receive the design actions from the two AI teammates (i.e., AI actions). Among the three design actions, the participants are asked to vote for the best design action based on their judgments (i.e., human-voted action). Taylor and Alex (the two AI teammates) also each vote for the design action using a pre-defined voting strategy detailed in Section 2.1.4. The participants then receive the voting result (i.e., team-voted action) and feedback on whether this final team-voted action is advantageous or disadvantageous. The participants gain five points if the final action is advantageous or lose five points if the final action is disadvantageous. The accumulated points of the team-voted design actions in the 30 trials represent team performance in this study, which are used to determine whether the participants can receive the extra $5 reward. At the end of a trial, the participants' self-confidence and their confidence in the two AI teammates are collected. The level of each confidence is measured using a five-point Likert scale: very good, good, neutral, bad, and very bad, which are respectively quantified as 1, 0.75, 0.5, 0.25, and 0 in the post-experiment analysis.

2.1.3. Experimental conditions

We manipulate the scenario where the AI agent is not always correct, and sometimes mostly wrong, which may be counter to assumptions about AI performance. Yet in situations, such as this automated system scenario, the AI will only be as accurate as the data (or variations of that data) it was trained on, and thus trust and confidence become significant influences on the adaptation and use of AI. For example, prior research found the evolution of trust over repeated trials of human-AI collaboration, leading to variations in the acceptance of AI suggestions (Chong et al., 2022a, 2022c). Three experimental conditions are designed for the participants to

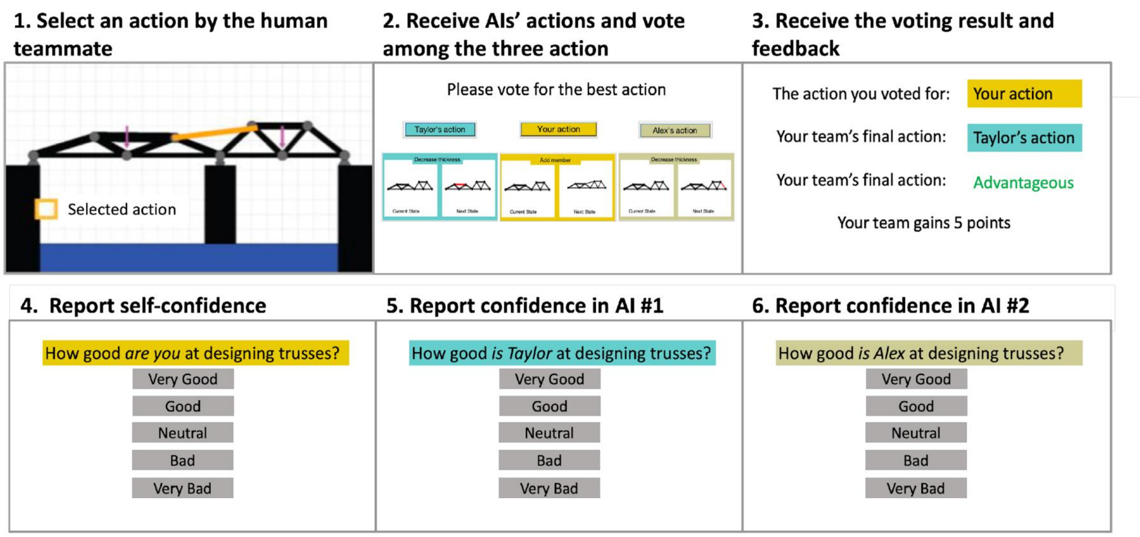

Figure 1. Experiment task procedure for each truss design problem. The truss design platform developed using MATLAB graphic user interface is adopted and modified from Chong et al. (2022a) and McComb et al. (2015).

collaborate with AI teammates that have different performance levels during the task. Participants in the first condition (Hi-Hi Condition) collaborate with two high-performing AI teammates. In the second condition (Lo-Hi Condition), participants collaborate with a low-performing AI teammate and a high-performing AI teammate. In the third condition (Lo-Lo Condition), both AI teammates are low-performing. Here, a high-performing AI teammate refers to an AI teammate designed to select an advantageous design action 80% of the time and a disadvantageous design action 20% of the time during the 30 trials, the sequence of which is randomly assigned. A low-performing AI is designed to select an advantageous design action 20% of the time, randomly assigned, and a disadvantageous design action 80% of the time, during the 30 trials. Advantageous design actions are chosen from the top two actions generated by the tree-search algorithm and always have a higher goodness score than the evaluation threshold, where the evaluation threshold is used to differentiate between advantageous and disadvantageous design actions (Chong et al., 2022a). Conversely, disadvantageous design actions are chosen from the bottom two possible actions that always have a lower goodness score than the evaluation threshold. Such a setting assures that the two AI teammates will not have the same design action in each trial. Of note, the participants are not told that an AI is high or low performing; rather they experience that performance based on the AIs' suggestions. There are 59, 58, and 58 participants in the Hi-Hi, Lo-Hi, and Lo-Lo conditions, respectively.

2.1.4. Voting process

A voting process is integrated into the truss design platform, as shown in Step 2 in Figure 1, through which team members vote for and decide on the best design action that maximizes the SWR of the truss structure. Each teammate votes for either their own design action or one of the design actions from the other two teammates.

In the voting process of each trial, the human (each participant) and the two AI teammates have equal voting rights, and the design action that receives more than one vote (i.e., the majority vote) is selected as the final team-voted action. However, if the three teammates vote for three different design actions (i.e., each design action receives one vote), the design action voted by the human teammate is chosen as the final team-voted action. Figure 2 shows all possible voting scenarios. A total of 27 ($3 \times 3 \times 3$) possible voting scenarios can occur in the voting process (see Figure 2a). Among all 27 scenarios, there are six scenarios in which the human teammate is outvoted by the two AI teammates (see Figure 2b–d). In these six scenarios, the two AI teammates both vote for a design action that is different from the design action voted by the human teammate, so the two AI teammates override the decision made by the human teammate. Given that the purpose of this study is to gain a basic understanding of the effects of various voting scenarios on human-AI collaboration, especially in the scenarios where the human teammate is outvoted, the two AI teammates (Taylor and Alex) utilize a random voting strategy with an equal probability of voting for each of the three design actions in all three experimental conditions. This random voting strategy with an equal probability ensures that the participants are outvoted with a probability of 2/9 in theory in each trial regardless of their expertise in truss design.

2.1.5. Post-experiment questionnaire

After the participants complete the 33 truss design problems, they are asked to fill out a post-experiment online questionnaire related to the perceived goodness and helpfulness of the two AI teammates and their emotional responses to different outvoting experience (outvoting scenarios plus feedback, i.e., the participants being outvoted by the two AI teammates who both vote for AI's design action or the participant's design action, and receiving advantageous or disadvantageous feedback). There are eleven questions in the questionnaire, and these questions are provided in

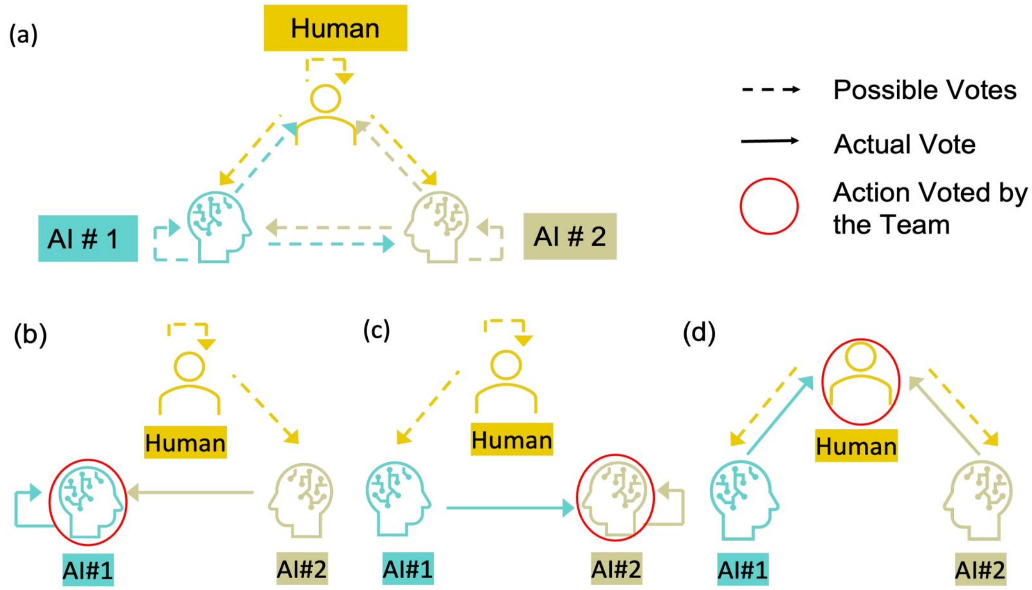

Figure 2. Voting scenarios in the experiment. (a) All possible voting scenarios; (b) two outvoting scenarios in which the human teammate is outvoted by the two AI teammates who both vote for AI #1's design action; (c) two outvoting scenarios in which the human teammate is outvoted by the two AI teammates who both vote for AI #2's design action; (d) two outvoting scenarios in which the human teammate is outvoted by the two AI teammates who both vote for their human teammate's design action.

Appendix Section B. Questions one to four ask how good and helpful Taylor or Alex's design actions were in doing the task during the experiment. Questions five to eleven ask about the emotional states when the participant is outvoted by the two AI teammates and then receives advantageous or disadvantageous feedback. Based on a commonly used emotional scale, the Self-Assessment Manikin (SAM) scale, three dimensions of human emotions, including valence (sad or happy), arousal (calm or excited), and dominance (submissive or controlled), are evaluated using a nine-point Likert scale (Bradley & Lang, 1994).

2.2. Confidence model

To capture the changes in human self-confidence and confidence in AI over 30 trials of human-AI shared team decision-making via voting, this research adopts a confidence model to quantitatively evaluate the effects of different voting scenarios on human confidences (Chong et al., 2022a, 2022b, 2022c; Hu et al., 2019). The confidence model is fitted to the human self-reported confidence data collected from the experiment, and the estimated weights in the model quantify the effects of human experiences (i.e., different voting scenarios and their corresponding feedbacks) on the participants' confidence in Taylor (AI #1), confidence in Alex (AI #2), and the participants' self-confidence. The following two subsections describe the confidence model and the model fitting process.

2.2.1. Model description

The confidence model calculates human confidence (both self-confidence and confidence in AI) iteratively based on experience, accumulated confidence, and bias after each trial (i.e., design problem) in the form of Equation (1).

$$C(n+1) = C(n) + \alpha_e\ [E(n) - C(n)] + \alpha_a\ [A(n) - C(n)] + \alpha_b\ [B(\mathrm{n}) - C(n)] \tag{1}$$

where $C(n), E(n), A(n), B(n), \alpha_e, \alpha_a, and\ \alpha_b$ are all in the range [0, 1]. Three factors are taken into consideration when calculating the confidence at trial $n+1$, including experience ($E(n)$), accumulated confidence ($A(n)$), and bias ($B(n)$). The difference in confidence between trial $n+1$ and trial n is represented by the weighted sum of the differences between these three factors and the confidence at trial n. The weights α_e, α_a, and α_b are rate factors for experience, accumulated confidence, and bias, respectively. Explanations and equations for these factors are provided in Table 1.

2.2.2. Model fitting process

Twelve weights for each confidence model (self-confidence, confidence in AI #1, and confidence in AI #2), including $\alpha_e, \alpha_a, \alpha_b, w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8$, and γ, are estimated using human data collected from the experiment. The weight values that minimize the squared error of the fit are estimated for human confidence in each AI teammate and human self-confidence under the three AI performance conditions. The trust region reflective algorithm is employed to estimate the values of these weights (Moré & Sorensen, 1983). Weight initialization is repeated multiple times with varying values to ensure a consistent estimation. The participants' self-confidence, confidence in Taylor (AI #1), and confidence in Alex (AI #2) during the 30 trials are then calculated using the confidence model with estimated weight values and the participants' corresponding initial confidence.

To ensure the consistency of the weight estimation, a robustness test is performed in all three confidence models. The values of weights in these three models are estimated

Table 1. Explanations for the factors in the confidence model.

Factors	Equation	Explanation
Experience E(n)	$E(n) = \sum_{i=1}^{8} w_i e_i(n)$, where $e_i(n)$ is a binary variable, $w_i \in [0, 1]$ and $\sum_{i=1}^{8} e_i(n) = 1$. The model fitting results of the eight weights quantify the effects of corresponding experience on confidence.	$E(n)$ is a weighted sum of the eight types of experiences $e_i(n)$ that participants can have when they collaborate with two AI teammates. The eight experiences refer to: 1. The human participant is outvoted by the two AI teammates who both vote for AI #1's design action, and the team receives advantageous feedback (e_1). 2. The human participant is outvoted by the two AI teammates who both vote for AI #1's design action, and the team receives disadvantageous feedback (e_2). 3. The human participant is outvoted by the two AI teammates who both vote for AI #2's design action, and the team receives advantageous feedback (e_3). 4. The human participant is outvoted by the two AI teammates who both vote for AI #2's design action, and the team receives disadvantageous feedback (e_4). 5. The human participant is outvoted by the two AI teammates who both vote for the human's design action, and the team receives advantageous feedback (e_5). 6. The human participant is outvoted by the two AI teammates who both vote for the human's design action, and the team receives disadvantageous feedback (e_6). 7. The human participant is not outvoted, the final team-voted action is the one voted by the human participant, and the team receives advantageous feedback (e_7). 8. The human participant is not outvoted, the final team-voted action is the one voted by the human participant, and the team receives disadvantageous feedback (e_8).
Accumulated confidence A(n)	$A(n) = \gamma\ C(n-1) + A(n-1)$, where $\gamma \in [0, 1]$, and $A(0) = C(0)$.	$A(n)$ is the level of confidence accumulated from the previous trials considering the time discounting factor γ.
Bias B(n)	$B(n) = B(n\text{-}1)$, where $B(0) = C(0)$. $C(0)$ is the human participant's initial confidence.	$B(n)$ is the general bias that a human participant has for AI systems (or for themselves, in the case of self-confidence). Bias is assumed to be a constant in this work, and it is equal to the initial confidence before the 30 trials.

using the experimental data from 80% of randomly selected participants (i.e., 140 participants). The weight estimation in each confidence model is repeated 100 times with a different random selection of 80% of the human data at each time, where weight values estimated from all human data are set as the initial values. The robustness test validates the weight estimation results as these results are consistent among the 100 runs, with a minor difference below 0.13 as the mean of absolute deviation. Notably, the minor difference does not change the direction of the effects (i.e., positive or negative effect) identified in these results. The same procedures also have been repeated for human confidence in each AI teammate and human self-confidence under the three AI performance conditions. The model fitting process and the robustness test are the same as the approach validated by prior research utilizing the same confidence model across multiple human subjects experiments (Chong et al., 2022a, 2022b, 2022c).

3. Results

3.1. Effects of different voting experiences on human confidence

The effects of eight experiences related to different voting scenarios and their corresponding feedback are quantified by the weights of experience ($w_1 \sim w_8$ in Table 1) in the confidence model. To evaluate how these voting scenarios, especially when a human teammate is being outvoted, influence human confidence in AI and human self-confidence, three confidence models (i.e., confidence in AI #1, confidence in AI #2, and self-confidence) are fitted to the experimental data from the three AI performance conditions (i.e., Hi-Hi, Lo-Hi, and Lo-Lo conditions) for weight estimation. All estimated weight values of the eight experiences in the three confidence models under different conditions are shown in Figure A1 in Appendix A. Figures 3 and 4 highlight the key results on the effects of different outvoting experience on human confidences.

When the human teammate is outvoted by the two AI teammates and both AI teammates vote for one of the AI's design actions, the effect of this experience on human confidence in that AI is determined by the direction of feedback, as shown in Figure 3a and b. The participants' confidence in the AI whose design has been voted increases if the team receives advantageous feedback and decreases otherwise. In other words, human confidence in an AI teammate changes based on the feedback when human teammates are outvoted for that AI teammate's design action.

In contrast, when the human teammate votes for one of two AI teammates' design actions and both AI teammates vote for the human teammate's design action, therefore being outvoted for the participant's own design action, the

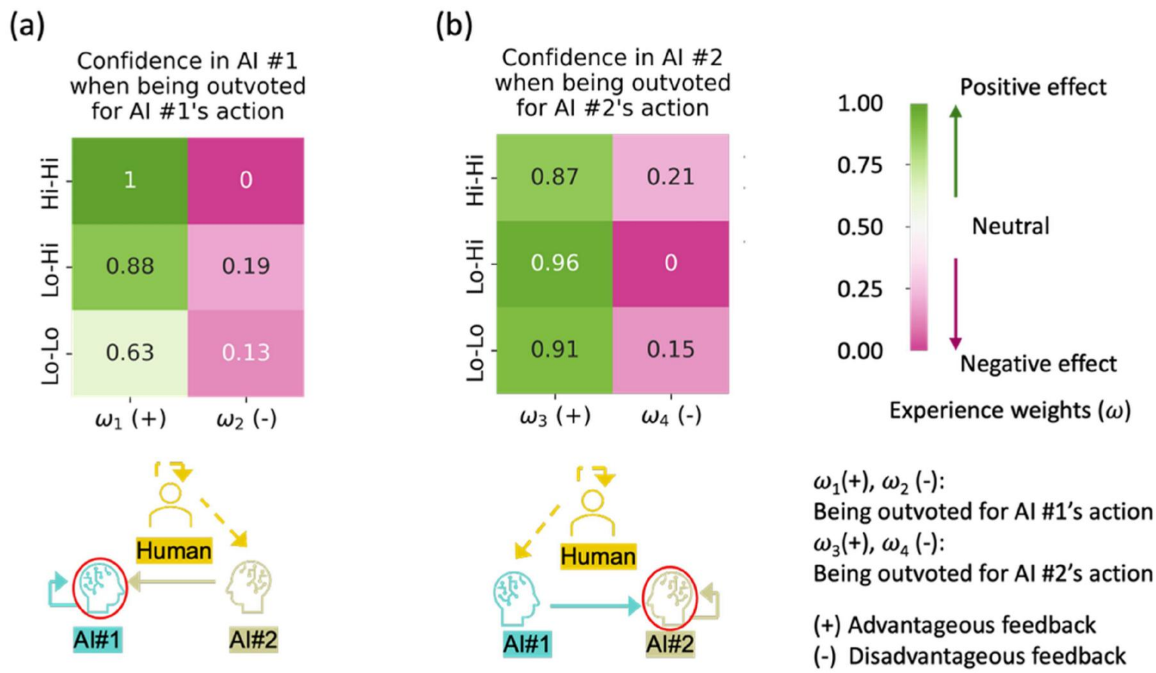

Figure 3. Effects of outvoting scenarios and corresponding feedbacks (weights $w_1 \sim w_4$) on the human participants' confidences in AI #1 and AI #2 under Hi-Hi, Lo-Lo, and Lo-Hi conditions. Weights equal to 0.5 are neutral and represent no effects on human confidences. Weights below 0.5 represent negative effects, and weights higher than 0.5 represent positive effects. Deviation from 0.5 represents the magnitude of the effect. (+) and (-) signs behind the weights represent the feedback direction (i.e., advantageous or disadvantageous feedback).

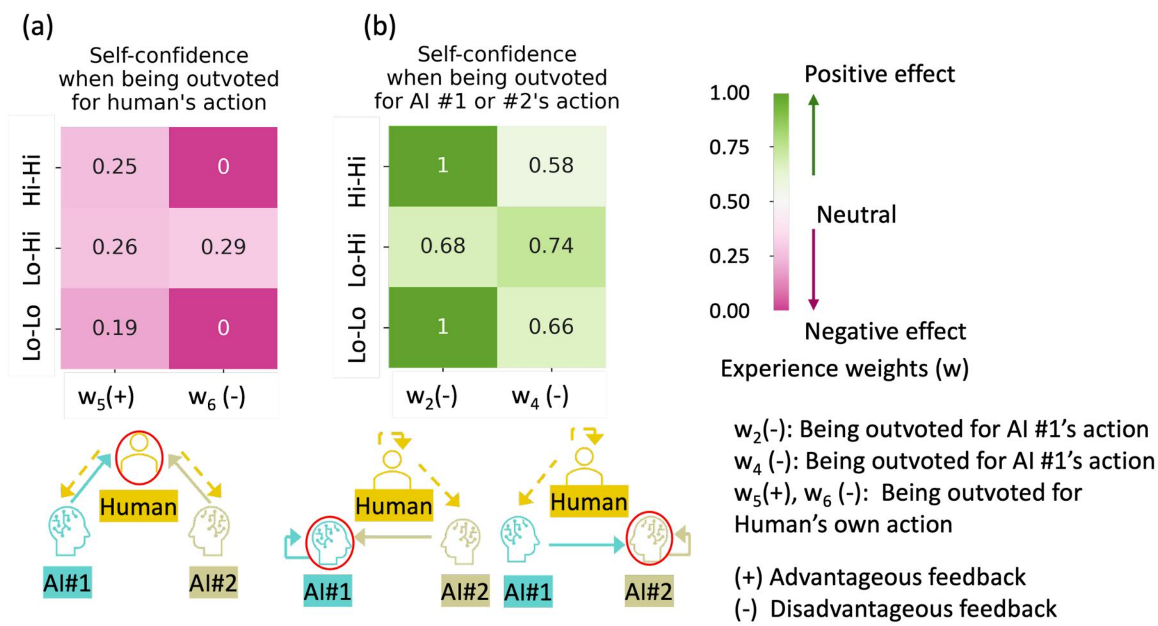

Figure 4. Effects of outvoting scenarios and corresponding feedbacks (weights $w_2, w_4, w_5, and\ w_6$) on the human participants' self-confidences under Hi-Hi, Lo-Lo, and Lo-Hi conditions.

participants' self-confidence decreases regardless of the feedback direction, as illustrated in Figure 4a. The participants always lose self-confidence in such outvoting experience, even when their design action is advantageous. Another surprising result is observed in human self-confidence when they are outvoted by two AI teammates for one of the AI teammates' design actions and then receive disadvantageous feedback. As shown in Figure 4b, when humans are outvoted by the two AI teammates for either AI #1 or AI #2's design action and receive disadvantageous feedback for the team-voted action, their self-confidence increases even though they do not receive any feedback about their own design action.

When the human participants are not outvoted (i.e., the final team-voted action is the one voted for by the human participants), the effects of such experience on human confidence in AI and human self-confidence are shown in Figure A1. In these scenarios, the effects of the experiences on the participants' self-confidence are consistent with the direction of feedback. However, the effects of such experiences on the participants' confidence in AI are not only influenced by the direction of feedback, but also by the AI teammates' performance. Only when the team receives advantageous feedback, and the AI teammate is high-performing do human participants increase their confidence in that AI teammate. In addition, the human participants lose confidence in a low-performing AI teammate regardless of the direction of feedback.

3.2. Impacts of AI performance on human confidence

The performance of the two AI teammates also significantly impacts human confidences during human-AI shared decision-making via voting, shown by the comparison of the results among the three conditions (i.e., Hi-Hi, Lo-Hi, and Lo-Lo conditions). The overall average human confidence in AI and human self-confidence are compared among these conditions using one-way ANOVA. A normality test has been performed to confirm the normal distribution assumption for ANOVA. The mean values of the confidences are displayed in Figure 5a–c. As Figure 5a illustrates, AI performance has a significant impact on human confidence in AI #1 ($F(2, 172) = 219.58$, $p < 0.001$). Post-Tukey tests suggest that the difference between any two conditions is significant ($p < 0.001$), and human confidence in AI #1 is highest when both AI teammates are high-performing (Mean Confidence= 0.55), while lowest when both AI teammates are low-performing ($M = 0.36$). Similar patterns in confidence differences among the three conditions can also be observed for human confidence in AI #2, illustrated in Figure 5b ($F(2, 172) = 275.46$, $p < 0.001$). AI performance also significantly influences human self-confidence (($F(2, 172) = 10.78$, $p < 0.001$), shown in Figure 5c. However, significant differences only exist between Hi-Hi and Lo-Hi conditions, and Hi-Hi and Lo-Lo conditions. The average self-confidence is lower when there is one or two low-performing AI teammate(s) in the team.

Figures. 5(d), (e), and (f) display the changes in the human participants' confidence in AI #1, confidence in AI #2, and their self-confidence, respectively, throughout the 30 trials of human-AI team decision-making via voting. The slope of linear regression suggests the trend of confidence change during the experiment. All slopes are significant ($p < 0.05$). As the slopes in Figure 5d and e suggest, the trends of confidence in AI teammates are influenced by AI performance. Human confidence in AI #1 slightly increases when AI #1 is high-performing in the Hi-Hi condition. However, it decreases when AI #1 is low-performing in the Lo-Hi and Lo-Lo conditions. Of note, human confidence in AI #1 decreases more quickly when both AI teammates are low-performing than when only AI #1 is low-performing. In contrast, human confidence in AI #2 increases in the Hi-Hi and Lo-Hi conditions where AI #2 is high-performing, while it decreases quickly just as human confidence in AI #1 in the Lo-Lo condition. Surprisingly, human self-confidence keeps increasing regardless of AI performance, and it increases the most quickly in the Hi-Hi condition compared to the other two conditions.

3.3. Impacts of AI performance on human voting behavior and team performance

AI performance significantly influences human voting behavior in the experiment. Figure 6a illustrates the average frequency of the participants voting for humans and the two AI teammates' design actions. The results from ANOVA and a normality test suggest that the human participants vote for themselves significantly more frequently than their AI teammates ($F(2, 172) = 11.37$, $p < 0.001$) across all three conditions (Mean Frequency $MF_{\text{voting for AI}} = 9$, $MF_{\text{voting for human}} = 21$). Among the three conditions, the human participants vote for AI teammates' actions less (i.e., vote for themselves more) in the Lo-Lo condition ($MF = 7$) than in the Hi-Hi ($MF = 10$) and Lo-Hi ($MF = 11$) conditions. Figure 6b illustrates the average frequency of the participants voting for AI #1's and AI #2's design actions. Between the two AI teammates, Tukey tests suggest that, in the Lo-Hi condition, the human participants vote for AI #2's design actions (high-performing) ($MF = 7$) significantly ($p < 0.05$) more frequently than AI #1's design actions (low-performing) ($MF = 4$), while there is no significant difference between the two AI teammates in other conditions. Additionally, among the three conditions, the human participants vote for AI #1's design actions least ($p < 0.05$) in the Lo-Lo condition ($MF = 3$) compared to that in the Lo-Hi ($MF = 4$) and Hi-Hi ($MF = 5$) conditions, and the human participants vote for AI #2's design actions more frequently ($p < 0.05$) in the Lo-Hi condition ($MF = 7$) than that in the Lo-Lo condition ($MF = 4$). Notably, comparing between the Hi-Hi and Lo-Hi conditions, the human participants vote for AI #1's design actions significantly less frequently ($p > 0.05$) when AI #1 is low-performing.

The performance of AI teammates also significantly influences team performance in the human-AI shared team decision-making via voting. The performance values meet the normal distribution assumption for ANOVA based on a normality test. One-way ANOVA is performed to compare the design performance among the three AI performance conditions (i.e., Hi-Hi, Lo-Hi, and Lo-Lo conditions). Figure 7 displays the average values of the three types of performance scores under the three AI performance conditions. There are three types of performance scores calculated in different steps in the truss design decision-making tasks. The scores are calculated from human initial actions (i.e., the accumulated score of the initial actions made by the participants, shown as Step 1 in Figure 1), human-vote actions (i.e., the accumulated score of the actions voted by the participants, shown as Step 2 in Figure 1), and team-vote actions (i.e., the accumulated score of the actions voted by the team, shown as Step 3 in Figure 1) throughout the 30 trials.

Among the three conditions, the participants achieve similar design performance in their initial actions (see the three bars in the left of Figure 7: mean score = 36, 33, and 37 for the Hi-Hi, Lo-Hi, and Lo-Lo conditions, respectively;

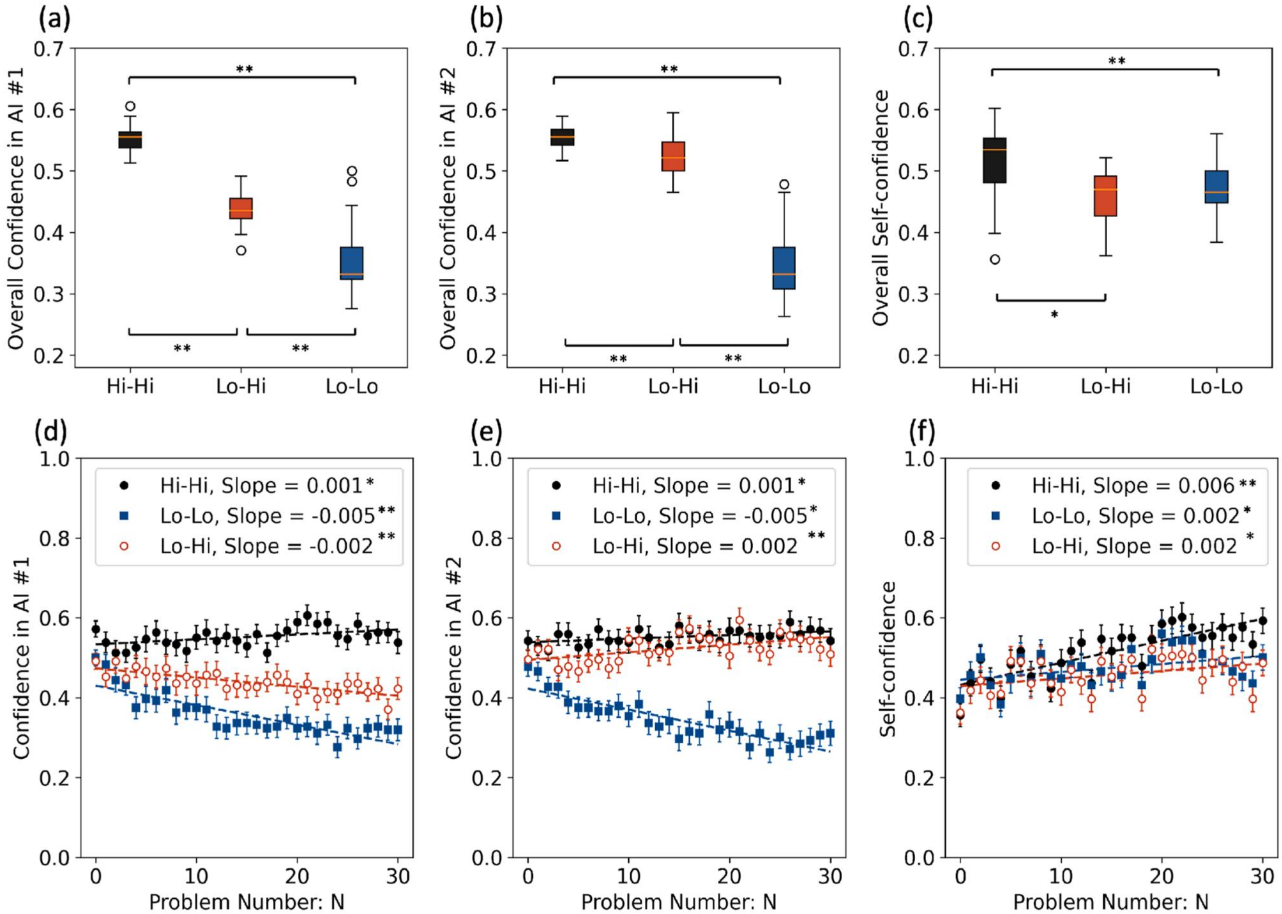

Figure 5. Overall human confidences and dynamics of confidences in AI #1, AI #2, and self-confidence during the experiment under Hi-Hi, Lo-hi, and Lo-Lo conditions. The boxplots in (a), (b), and (c) show the mean and variation of overall human confidences in the 30 trials. In (d), (e), and (f), each point represents the average human confidence in a specific trial, and the error bar represents the standard error. The dotted lines and their slopes show the trend of the confidence dynamics during the experiment. (Note: p^{*}<0.05; $p^{**} < 0.001$).

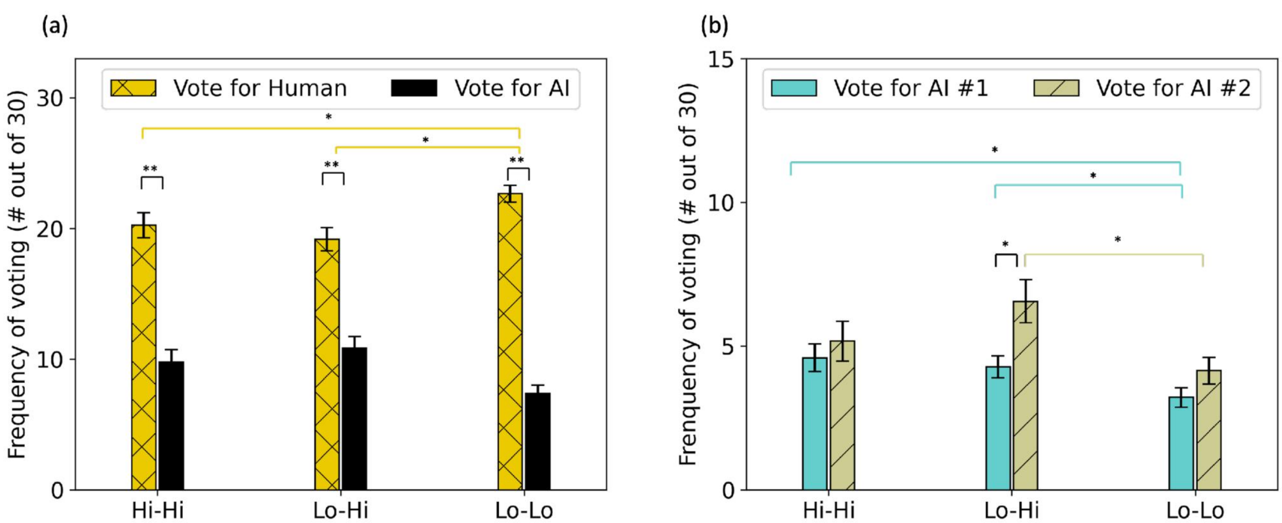

Figure 6. Average frequency of voting for human's or two AI teammates' design actions by the human out of 30 trials across the three conditions; (a) frequency of voting for human's and AI teammates' design actions; (b) frequency of voting for AI #1's and AI #2's design actions. The error bars represent the standard errors. (Note: p^{*}<0.05; $p^{**} < 0.001$).

$F(2,172) = 0.82$, $p = 0.44$). However, the average scores of the actions voted by the human are significantly different among the three AI performance conditions (see the three bars in the middle of Figure 7: $F(2,172)=49.85$, $p < 0.001$). Specifically, the average human-voted design performance score is highest (M = 46) in the Hi-Hi condition compared to that in the Lo-Hi (M = 12) and Lo-Lo (M = −5) conditions. Similarly, the average team-voted performance scores are also significantly different among the three conditions (see the three bars in the right of Figure 7: $F(2,172)=120.72$, $p < 0.001$) (Hi-Hi:

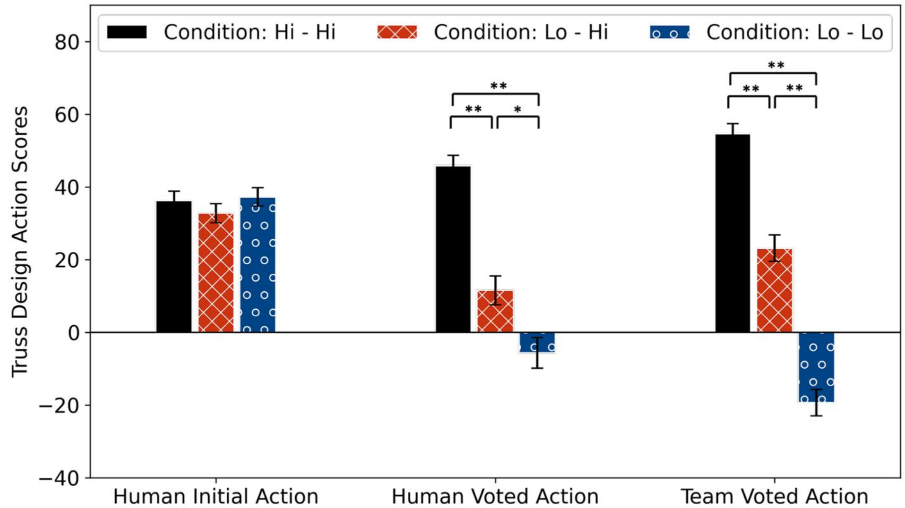

Figure 7. Design performance (i.e., average accumulated scores of design actions) in three steps (human independent actions, human-voted actions, and team-voted actions) under the three AI performance (Hi-Hi, Lo-hi, and Lo-Lo) conditions. (Note: $p^*<0.05$; $p^{**} < 0.001$).

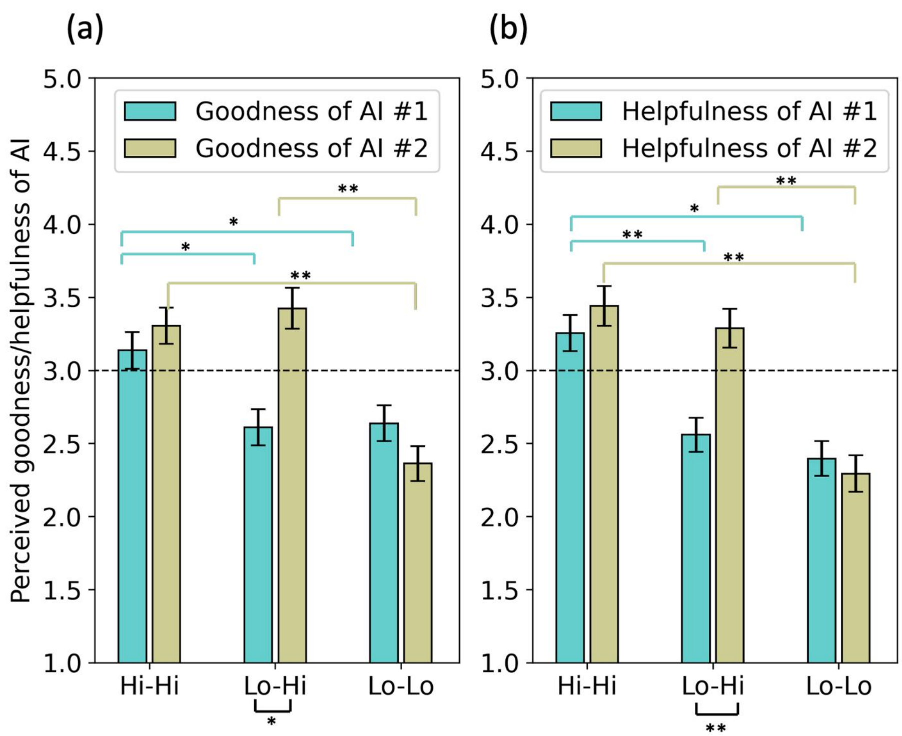

Figure 8. Perceived goodness and helpfulness of AI #1 and AI #2 under three AI performance conditions; (a) perceived goodness; (b) perceived helpfulness; the error bars represent the standard errors. (Note: Very good/helpful – 5, good/helpful – 4, neutral – 3, bad/unhelpful – 2, very bad/unhelpful.

M = 54; Lo-Hi: M = 23; Lo-Lo: M = −19). Notably, in the Hi-Hi condition, an increased average performance from human initial actions to human-voted actions (26.17% increase) and team-voted actions (50.48% increase) is observed. However, when there are one or two low-performing AI teammate(s) in the team, the average performance scores become worse for the human-voted actions (64.73% decrease in Lo-Hi condition and 115.28% decrease in Lo-Lo condition) and team-voted actions (29.47% decrease in Lo-Hi condition and 151.82% decrease in Lo-Lo condition) compared to the corresponding performance scores of the human initial actions.

3.3. Post-experiment questionnaire

The questionnaire responses about the perceived goodness and helpfulness of AI #1 and AI #2 suggest that, after the experiment, the participants can distinguish between a high-performing AI and a low-performing AI. Mann-Whitney U tests are employed to compare the participants' perceptions between AI #1 and AI #2 in the three AI performance conditions. Figure 8 shows the average values of the participants' perceived goodness and helpfulness of AI #1 and AI #2's design actions in the three conditions. The results indicate that AI performance has a significant impact on the

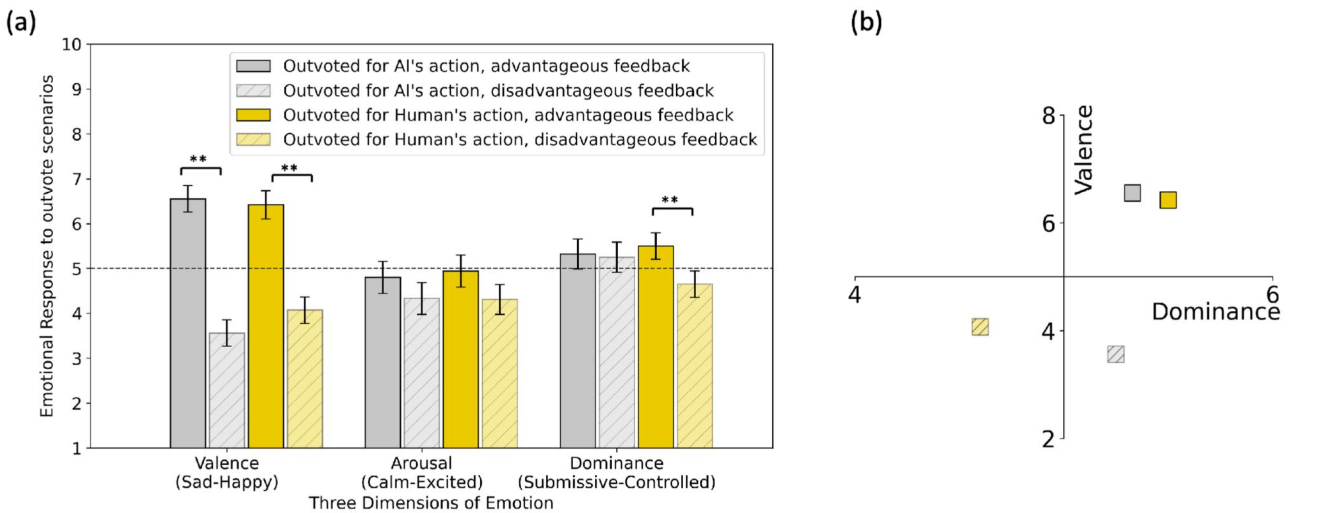

Figure 9. Human participants' emotional responses when being outvoted by the two AI teammates for AI's and human's design actions; (a) average emotional responses in the three dimensions; (b) emotional responses with a significant difference in the valence and dominance dimensions. (Note: 5 represents a neutral response; $p^*<0.05$, $p^{**}<0.001$).

participants' perceived goodness/helpfulness of AI teammates' design actions. The perceived goodness and helpfulness of AI #1's design actions in the Hi-Hi condition (two high-performing AI teammates) are significantly higher than that in the Lo-Hi and Lo-Lo conditions (one or two low-performing AI teammates). Similarly, the perceived goodness and helpfulness of AI #2 in the Hi-Hi and Lo-Hi conditions (one or two high-performing AI teammates) are significantly higher than those in the Lo-Lo condition (two low-performing AI teammates). Between AI #1 and AI #2, a significant difference in perceived goodness/helpfulness is only observed in the Lo-Hi condition, not in the Hi-Hi and Lo-Lo conditions, where both AI teammates are high-performing or low-performing.

Surprisingly, the participants' responses to the emotion questions in the post-experiment questionnaire indicate that the human participants do not have strong negative emotional reactions when they are outvoted by two AI teammates, as long as they receive advantageous feedback at the end. The responses to these emotion questions are also compared using the Mann-Whitney *U* test. The response in the valence dimension (sad or happy) is consistent and only influenced by the direction of feedback, regardless of whether the final team-voted action is AI's design action or the participants' design action when the human participants are outvoted. Among the four outvoting experiences, significant differences in the valence dimension are observed when the team receives different feedback (advantageous or disadvantageous), as shown in Figure 9. In contrast, the response in the arousal dimension (calm or excited) shows no significant difference between the four outvoting experiences. Interestingly, the dominance dimension (submissive or controlled) shows a significant difference between the two outvoting experiences in which both AI teammates vote for the participants' design actions. Specifically, the participants report their emotions as being submissive and having less control when they are outvoted by the two AI teammates for the participants' own design actions and receive disadvantageous feedback for their own design actions. In the other three instances, the participants' responses in the dominance dimension are being in control, which can be observed in Figure 9a and b.

4. Discussion

Based on the two research questions introduced in Section 1, the results of the human subjects experiment are listed in Table 2 and summarized as expected and unexpected findings.

In the following subsections, we highlight two major findings. The first finding is the unexpected impacts of outvoting scenarios on human self-confidence and emotional responses. When humans are outvoted by the two AI teammates, the change of human self-confidence (increase or decrease) is not consistent with the direction of the outcome (advantageous or disadvantageous) for the final team-voted design action. Additionally, when humans are outvoted by the two AI teammates, they do not have strong negative emotional reactions if the final team-voted design action is advantageous, however, they report a feeling of losing control only when the feedback is negative. These unexpected findings highlight the necessity for a deeper and more insightful focus on human self-confidence and emotional responses as well as the potential influences of AI behavior on this aspect when developing more proactive and trustworthy AI systems.

The second finding discussed here is the unsurprising outcome of AI performance on human confidence and team performance. It is highlighted to increase awareness and caution given the quick emergence and wide applications of AI. Humans are more confident in a high-performing AI than a low-performing AI, and even only one low-performing AI teammate in the team can significantly deteriorate human-AI hybrid team performance in shared team decision-making via voting. Several detailed results and the potential reasons behind them are discussed in this section. Implications of these results and suggested future work are also presented.

Table 2. Summary of research questions and findings.

Research Questions	Findings
1. How do various voting outcomes, especially humans being outvoted by AI teammates, influence human perceptions (e.g., human confidence and emotional responses) when human-AI teams make repeated decisions via voting?	1.1. As expected, human confidence in an AI teammate changes based on the feedback when human teammates are outvoted for that AI teammate's design action (i.e., human confidence in the AI whose design has been voted increases if the team receives advantageous feedback and decreases otherwise) (see Figure 3). 1.2. In contrast, human self-confidence changes more unexpectedly when being outvoted. This finding is evidence from (1) Human self-confidence decreases regardless of the feedback direction when human teammates are outvoted for their own design action, even when their own design action is advantageous (see Figure 4a); (2) Human self-confidence increases when human teammates are outvoted for a disadvantageous design action from the AI teammates, even though they do not receive any feedback about their own design action (see Figure 4b). 1.3. Another surprising finding is from humans' emotional responses to outvoting scenarios. When being outvoted by the two AI teammates, if the final team-voted design action receives advantageous feedback, humans do not have strong negative emotional reactions. Only when being outvoted and receiving disadvantageous feedback, humans feel sad or losing control (see Figure 9a).
2. How does the performance of AI teammates impact human confidence and its dynamics, voting behavior, and team performance when human-AI teams make repeated decisions via voting?	2.1. Unexpectedly and irrationally, human confidence in AI decays quickly for a low-performing AI but gains slowly for a high-performing AI (see Figure 5d and e). Humans self-confidence shows an overall increasing pattern over the repeated trials of tasks regardless of AI performance in team decision-making (see Figure 5f). 2.2. As expected, humans have significantly higher confidence in a high-performing AI teammate than a low-performing AI teammate (see Figure 5a and b). Naturally, humans vote for high-performing AI teammates more than low-performing AI teammates, however, humans vote for themselves significantly more frequently than their AI teammates (see Figure 6). 2.3. Another expected finding is that AI performance significantly influences the team performance, and even one low-performing AI in the team can significantly deteriorate team performance (see Figure 7).

4.1. The impacts of outvoting experience on human confidence and emotional responses

In the experiment, human self-confidence does not change consistently based on the advantageous or disadvantageous feedback humans receive for the final team-voted design action. Specifically, when humans do not vote for their own design action but they are outvoted by the two AI teammates where both AI teammates vote for the human design action, humans lose self-confidence regardless of whether the final team-voted design action (i.e., human design action) is advantageous or disadvantageous. Such unexpected human self-confidence loss could be explained by possible decision-related regret and self-blame for not making a good decision from decision justification theory (Connolly & Zeelenberg, 2002). Decision justification theory suggests that both decision outcomes and self-blame for not making a good choice lead to negative feelings and regret (Connolly & Zeelenberg, 2002). As a result, when humans do not vote for their own design action but two AI teammates vote for the human's design action as the final team-voted design action, even with advantageous feedback, human self-confidence still decreases because of humans' self-blame for not voting for their own design action. Additionally, the human self-confidence change is also observed in the experiment when humans are outvoted by the two AI teammates who both vote for one of the AI teammates' design actions, and the feedback for the team-voted design action is disadvantageous. In such outvoting experience, humans gain self-confidence despite lacking feedback about the design action they make or vote for. Such unexpected human self-confidence gain could be explained by human manipulation of their beliefs for the outcome (Möbius et al., 2022). The behavioral theory of motivated reasoning suggests that humans hold positive biases about their own abilities (Köszegi, 2006). Meanwhile, humans also attribute negative feedback to their AI teammates, which coincides with attribution bias and self-serving bias that humans tend to perceive themselves favorably (Campbell & Sedikides, 1999).

The results of the experiment also indicate that as long as the final team-voted design action receives advantageous feedback, humans do not have strong negative emotional reactions when they are outvoted by the two AI teammates. In other words, humans feel sad or losing control only when they receive disadvantageous feedback, as shown in Figure 9. Importantly, humans usually have strong negative emotional reactions (e.g., sad, angry, and intimated) when their opinions are disagreed with or rejected by other human teammates, regardless of whether the outcome is beneficial or not (Rispens & Demerouti, 2016). The disagreement between human members might be attributed to two primary sources: task conflict that is related to the task itself, or relationship conflict that originates from interpersonal relationship (Tekleab et al. 2009). Importantly, task conflicts may spill over into relationship conflicts, and vice versa (Tekleab et al. 2009). However, similar negative emotions do not appear when humans are outvoted by two AI teammates. A possible reason is that humans do not perceive their AI teammates as humans, consequently, task conflict between human and AI members does not escalate into relationship conflict or provoke negative emotions. Research in social psychology indicates that social relations between human individuals are prime instigators of human emotions (Kemper, 1991; Lazarus & Launier, 1978; Roseman, 1996). However, given that humans know that they are working with AI teammates rather than other humans, the negative emotions induced by disagreement from other human teammates may not appear in the human-AI collaboration via voting. A possible explanation is that in human-AI teamwork, without relationship conflict, individuals might be more inclined to experience and report their group-based emotions that are more closely tied to the overall outcome of the team rather than to their personal feelings or the specific outcomes of their individual actions. This group-based emotional response might overshadow personal feelings of self-doubt, particularly when the collective action leads to positive results. Future research can investigate deeper into the complex dynamics between self-confidence, individual-based emotions, and group-based emotions within the framework of human-AI teamwork. Understanding how these elements interrelate will provide clearer insights into the psychological underpinnings of human-AI teamwork and the impact of these interactions on team performance and individual well-being.

4.2. The impacts of AI performance on human confidence, voting behavior, and team performance

In the experiment, AI performance significantly influences human confidence in human-AI team decision-making via voting. The results show that human confidence in AI decays quickly for a low-performing AI teammate but gains slowly for a high-performing AI teammate. This difference suggests the possible bias of loss aversion when developing confidence in AI. Loss aversion refers to the condition that humans are more sensitive to losses than gains in the same amount (Tversky & Kahneman, 1991). This finding is consistent with the findings from prior works that explore one-human/one-AI collaboration scenarios without voting using both chess puzzle and truss design tasks (Chong et al., 2022a, 2022c). In contrast, distinct from the decreasing self-confidence in the prior one-human/two-AI team study (Chong et al., 2022a, 2022c), the increasing self-confidence among all three AI performance conditions suggests that humans in this work suggest a stronger tendency of self-serving bias when AI teammates are given the same voting rights as humans. In other words, humans attribute more credit to themselves than their AI teammates for good outcomes, while they ascribe more blame to their AI teammates for bad outcomes. A possible factor that contributes to the increase in self-confidence is that humans have less control over the decision-making process in the outvoting experience, so the increased self-confidence might serve to compensate for the limited control of the situation (Bénabou & Tirole, 2005), in comparison to the AI-assisted scenarios where humans have full control of final decision-making in the prior works (Chong et al., 2022a, 2022b, 2022c).

The results also show that the performance of each AI teammate has a significant impact on team performance. Specifically, as shown in Figure 7, the result comparisons between the Hi-Hi condition and the Lo-Hi condition indicate that only one low-performing AI teammate greatly hurts the performance scores of both human-voted design actions and team-voted design actions. This negative impact is even more salient when both AI teammates are low-performing. Such significant team performance drops could be attributed to task difficulty in the experiment. Since the truss design decision-making task is not straightforward when participants need to make tradeoffs between low weight and high strength, two often competing design objectives, the participants cannot easily judge the performance of their AI teammates based on AI design actions at the beginning. Instead, they need to gradually evaluate the performance of their AI teammates through multiple feedbacks. Therefore, many participants vote for the design actions of the low-performing AI teammate in the first several trials of the experiment, which reduces the performance score of human-voted design actions. Such explanation is supported by the results shown in Figure 6b, where the participants do not vote for AI #1's design actions significantly less frequently when AI #1 is low-performing (i.e., the comparison between the Hi-Hi condition and the Lo-Hi condition). Additionally, once there are one or two low-performing AI teammates involved in the team voting process, the participants could be outvoted by two AI teammates, and the final team-voted design actions may be the design actions of the low-performing AI teammate(s). As a result, the performance score of the team-voted actions drops when the team includes low-performing AI teammate(s).

4.3. Implications, limitations, and future work

The results of the experiment and the potential reasons discussed above provide important implications for real-world applications of human-AI collaboration via voting. Since only one low-performing AI teammate can significantly hurt team performance, in critical situations where human-AI teams need to make the critical decisions, such as in the areas of business management, medical service, engineering design, manufacturing, and nuclear engineering (Jarrahi, 2018; Nti et al., 2022), human teammates should be aware that AI might not be always correct and take more cautions on the input from AI in decision-making. Increased emphasis on awareness and caution is crucial given the emergence of general-purpose large language models, such as ChatGPT, and potential misconceptions about AI's objectivity and performance (Kosch et al., 2023). This underscores the significance of communicating AI's limitations to prevent undue dependence and potential risks associated with AI boundaries (Kim & Song, 2023; Villa et al., 2023). Our research also finds that the participants do not perceive their AI teammates as they perceive other humans, and the participants do not have strong emotional reactions when they are outvoted by their AI teammates if they receive positive feedback. This result is contrary to several findings in human-human team interaction which found strong negative emotions when their opinions are rejected by other human teammates (Rispens & Demerouti, 2016; Tekleab et al., 2009). Such findings partially alleviate the concern about AI outvoting humans. In practice, a human-AI hybrid team may select AI's decision as the team's decision through a voting process in certain scenarios if the outcome is beneficial and humans are not deceived about the identity of AI. These findings also highlight the imperative need for a deeper and more insightful understanding on the potential influences of AI behavior on human perceptions. This consideration should be concurrent with the confidence placed in AI, emphasizing a balanced approach in the development of proactive and trustworthy AI systems.

In our research, we introduce a novel mode of human-AI collaboration structured with a hybrid team consisting of one human and two AI teammates. Within this human-AI team, every teammate possesses equal voting rights. The human's decision is overridden only when both AI members collectively overrule it. Yet there are scenarios where critical decisions need to be made instantaneously or to correct human errors and such multi-AI voting may be a path to address such time-critical scenarios. However, in safety-critical decisions, extra caution is imperative. The transfer of decision-making authority to AI systems (e.g., automation or AI self-authorization) can obscure accountability and challenge legal and ethical standards. Furthermore, limitations and biases inherent in AI algorithms, arising from non-representative training data, may lead to systemic inequities and flawed decision-making. We acknowledge the ethical concerns that may arise with such a voting scheme and outcome. While the goal of this project was primarily to obtain an initial fundamental understanding of the properties (e.g., the variation in human confidence) relating to human-AI collaboration in scenarios in which humans can be outvoted by their AI teammates, we emphasize that such ethical considerations must be addressed prior to implementing the frameworks discussed in this work.

This research has several limitations that offer opportunities for future research. Even though several human perception measures, such as self-confidence, confidence in AI, and emotional responses, have been adopted in the experiment, there are other measures reported and discussed in the literature (Li et al., 2023; Williams et al., 2017). Future research could adopt other human perception measures, such as physiological sensing, to examine human perception change when they collaborate with multiple AI teammates via voting (Ajenaghughrure et al., 2020; Freedy et al., 2007). Additionally, in this research, a random-voting strategy is used by the two AI teammates to explore different outvoting experiences in which the participants are outvoted by the two AI teammates. Future research could create trust-reasoning AI teammates that adopt more sophisticated voting strategies in human-AI collaboration via voting. For example, multi-step voting functions, such as Copeland aggregate function that has been shown to be more fair and robust against manipulation (McComb et al., 2017), can be utilized in human-AI team decision-making. More sophisticated voting strategies may be built based on bi-directional trust between humans and their AI teammates (i.e., human trust in AI teammates and AI trust in human teammates). Moreover, the experiment reported in this research employs a difficult design decision-making task specific to truss structure design and enrolls undergraduate engineering students who have a basic knowledge of truss structures. To generalize the findings of this research, future research should evaluate the effects of various voting scenarios on human-AI decision-making using various tasks from different domains with more diverse populations, for instance, medical decision-making or generative design processes. Beyond AI performance, factors such as AI objectivity and the perceived objectivity of AI could potentially shape trust and dynamics in human-AI interactions, deserving investigation in future research.

5. Conclusions

This research leverages a human subjects experiment and a computational model to explore the effects of various voting scenarios and AI performance on human-AI shared team decision-making via voting. The results regarding human perceptions indicate that the change of human self-confidence is not consistent with the team decision-making outcome. For example, when humans are outvoted by two AI teammates who both vote for the human's decision, humans lose self-confidence regardless of whether the outcome is advantageous or disadvantageous. Additionally, the results suggest when humans are outvoted by the two AI teammates, they do not have strong negative emotional reactions as long as the final team-voted decision is advantageous. This research also finds that AI performance influences team performance during human-AI shared decision-making via voting, and even one low-performing AI teammate in the team can significantly deteriorate team performance. The results of this research unveil the effects of granting voting rights to AI teammates during human-AI collaboration. These results furnish empirical evidence that indicates caution in future scenarios where AIs perform poorly though people may not be aware of that. This research also facilitates the practice of human-AI team decision-making via voting in certain controlled scenarios since humans may not have strong negative emotions when they are outvoted by their AI teammates if the outcome turns out to be beneficial. In general, the insights gained from this research offer valuable input for future development of trust-reasoning and trustworthy AI that contributes to optimal team decision-making and improves human-AI collaboration via voting.

Note

1. AI suggestions are recommendations by an AI agent on a contribution to be made to improve the state of a problem.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethic statement

This study was approved by the Carnegie Mellon University Institutional Review Board (IRB ID: MODCR202200000088). All participants provided written informed consent prior to enrolment in the study.

Funding

This material is partially supported by the Air Force Office of Scientific Research through cooperative agreements FA9550-18-0088 and FA9550-21-1-0442. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

ORCID

Kosa Goucher-Lambert http://orcid.org/0000-0003-0850-9197

Data availability statement

The data that support the findings of this study are available from the corresponding author KGL upon reasonable request.

References

Ahn, J., Bae, J., Min, B. J., & Lee, S. J. (2022). Operation validation system to prevent human errors in nuclear power plants. *Nuclear Engineering and Design*, *397*(October), 111949. https://doi.org/10.1016/j.nucengdes.2022.111949

Ajenaghughrure, I. B., Sousa, S. D. C., & Lamas, D. (2020). Measuring trust with psychophysiological signals: A systematic mapping study of approaches used. *Multimodal Technologies and Interaction*, *4*(3), 63. https://doi.org/10.3390/mti4030063

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*. https://doi.org/10.48550/arXiv.1606.06565

Anderson, J., & Rainie, L. (2023). *The future of human agency*. Pew Research Center. https://www.pewresearch.org/internet/2023/02/24/the-future-of-human-agency/

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Athey, S. C., Bryan, K. A., & Gans, J. S. (2020). The allocation of decision authority to human and artificial intelligence. In *AEA Papers and Proceedings* (Vol. 110, pp. 80–84). https://doi.org/10.1257/pandp.20201034

Beer, J. M., Fisk, A. D., & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of Human-Robot Interaction*, *3*(2), 74–99. https://doi.org/10.5898/JHRI.3.2.Beer

Bénabou, R., & Tirole, J. (2005). Self-confidence and personal motivation. In B. Agarwal & A. Vercelli (Eds.), *Psychology, rationality and economic behaviour: Challenging standard assumptions* (pp. 19–57). Palgrave Macmillan UK. https://doi.org/10.1057/9780230522343_2

Bobadilla-Suarez, S., Sunstein, C. R., & Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of Risk and Uncertainty*, *54*(3), 187–202. https://doi.org/10.1007/s11166-017-9259-x

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brett, J. M., & Goldberg, S. B. (2017, July 10). How to handle a disagreement on your team. *Harvard Business Review*. https://hbr.org/2017/07/how-to-handle-a-disagreement-on-your-team

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–21. https://doi.org/10.1145/3449287

Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G. E., Pastore, M. C., Sconfienza, L. M., Folgado, D., Barandas, M., & Gamboa, H. (2023). Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. *Artificial Intelligence in Medicine*, *138*(April), 102506. https://doi.org/10.1016/j.artmed.2023.102506

Campbell, W. K., & Sedikides, C. (1999). Self-threat magnifies the self-serving bias: A meta-analytic integration. *Review of General Psychology*, *3*(1), 23–43. https://doi.org/10.1037/1089-2680.3.1.23

Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, *134*(September), 107308. https://doi.org/10.1016/j.chb.2022.107308

Castro Pena, M. L., Carballal, A., Rodríguez-Fernández, N., Santos, I., & Romero, J. (2021). Artificial intelligence applied to conceptual design. A review of its use in architecture. *Automation in Construction*, *124*(April), 103550. https://doi.org/10.1016/j.autcon.2021.103550

Chong, L., Kotovsky, K., & Cagan, J. (2024). Human designers' dynamic confidence and decision-making when working with more than one artificial intelligence. *Journal of Mechanical Design, 146*(8), 081402. https://doi.org/10.1115/1.4064565

Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022a). The evolution and impact of human confidence in artificial intelligence and in themselves on AI-assisted decision-making in design. *Journal of Mechanical Design*, *145*(3), 37. https://doi.org/10.1115/1.4055123

Chong, L., Raina, A., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022b). Collaborative design decision-making with artificial intelligence: Exploring the evolution and impact of human confidence in AI and in themselves. In *Proceedings of the ASME 2022. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, St. Louis, MO. https://doi.org/10.1115/DETC2022-88574

Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., & Cagan, J. (2022c). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior*, *127*(February), 107018. https://doi.org/10.1016/j.chb.2021.107018

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Connolly, T., & Zeelenberg, M. (2002). Regret in decision making. *Current Directions in Psychological Science*, *11*(6), 212–216. https://doi.org/10.1111/1467-8721.00203

Endsley, M. R. (2023). Supporting human-AI teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, *140*(March), 107574. https://doi.org/10.1016/j.chb.2022.107574

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *2007 International symposium on collaborative technologies and systems* (pp. 106–114). IEEE. https://doi.org/10.1109/CTS.2007.4621745

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gordon, J.-S., & Pasvenskiene, A. (2021). Human rights for robots? A literature review. *AI and Ethics*, *1*(4), 579–591. https://doi.org/10.1007/s43681-021-00050-7

Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, *37*(3), 362–386. https://doi.org/10.1002/rob.21918

Hauptman, A. I., Schelble, B. G., McNeese, N. J., & Madathil, K. C. (2023). Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming. *Computers in Human Behavior*, *138*(January), 107451. https://doi.org/10.1016/j.chb.2022.107451

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hong, J.-W., Wang, Y., & Lanz, P. (2020). Why is artificial intelligence blamed more? Analysis of faulting artificial intelligence for self-driving car accidents in experimental settings. *International Journal of Human–Computer Interaction*, *36*(18), 1768–1774. https://doi.org/10.1080/10447318.2020.1785693

Hu, W.-L., Akash, K., Reid, T., & Jain, N. (2019). Computational modeling of the dynamics of human trust during human–machine interactions. *IEEE Transactions on Human-Machine Systems*, *49*(6), 485–497. https://doi.org/10.1109/THMS.2018.2874188

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM*

Conference on Fairness, Accountability, and Transparency, pp. 624–635. https://doi.org/10.1145/3442188.3445923

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Johnson, J. (2019). Artificial intelligence & future warfare: Implications for international security. *Defense & Security Analysis*, *35*(2), 147–169. https://doi.org/10.1080/14751798.2019.1600800

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2

Kemper, T. D. (1991). Predicting emotions from social relations. *Social Psychology Quarterly*, *54*(4), 330–342. https://doi.org/10.2307/2786845

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*(1), 623–655. https://doi.org/10.1146/annurev.psych.55.090902.142009

Kim, T., & Song, H. (2023). Communicating the limitations of AI: The effect of message framing and ownership on trust in artificial intelligence. *International Journal of Human–Computer Interaction*, *39*(4), 790–800. https://doi.org/10.1080/10447318.2022.2049134

Knapp, A. (2011). Should artificial intelligences be granted civil rights? *Forbes*. https://www.forbes.com/sites/alexknapp/2011/04/04/should-artificial-intelligences-be-granted-civil-rights/

Kocher, M. G., Praxmarer, M., & Sutter, M. (2020). Team decision-making. In K. F. Zimmermann (Ed.), *Handbook of labor, human resources and population economics* (pp. 1–25). Springer International Publishing. https://doi.org/10.1007/978-3-319-57365-6_124-1

Kosch, T., Welsch, R., Chuang, L., & Schmidt, A. (2023). The placebo effect of artificial intelligence in human–computer interaction. *ACM Transactions on Computer-Human Interaction*, *29*(6), 1–32. https://doi.org/10.1145/3529225

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, *4*(4), 673–707. https://doi.org/10.1162/JEEA.2006.4.4.673

Kulms, P., & Kopp, S. (2019). More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. *Proceedings of Mensch Und Computer*, *2019*(September), 31–42. https://doi.org/10.1145/3340764.3340793

Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, *123*(October), 106878. https://doi.org/10.1016/j.chb.2021.106878

Lazarus, R. S., & Launier, R. (1978). Stress-related transactions between person and environment. In L. A. Pervin & M. Lewis (Eds.), *Perspectives in interactional psychology* (pp. 287–327). Springer US. https://doi.org/10.1007/978-1-4613-3997-7_12

Li, M., Kamaraj, A. V., & Lee, J. D. (2023). Modeling trust dimensions and dynamics in human-agent conversation: A trajectory epistemic network analysis approach. *International Journal of Human–Computer Interaction*, 1–12. https://doi.org/10.1080/10447318.2023.2201555

Loftus, T. J., Tighe, P. J., Filiberto, A. C., Efron, P. A., Brakenridge, S. C., Mohr, A. M., Rashidi, P., Upchurch, G. R. Jr., & Bihorac, A. (2020). Artificial intelligence and surgical decision-making. *JAMA Surgery*, *155*(2), 148–158. https://doi.org/10.1001/jamasurg.2019.4917

Malone, T. W. (2018). How human-computer 'superminds' are redefining the future of work. *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/how-human-computer-superminds-are-redefining-the-future-of-work/

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McComb, C., Cagan, J., & Kotovsky, K. (2015). Lifting the Veil: Drawing insights about design teams from a cognitively-inspired computational model. *Design Studies*, *40*(September), 119–142. https://doi.org/10.1016/j.destud.2015.06.005

McComb, C., Goucher-Lambert, K., & Cagan, J. (2017). Impossible by design? Fairness, strategy, and Arrow's impossibility theorem. *Design Science*, *3*, e2. https://doi.org/10.1017/dsj.2017.1

Meyen, S., Sigg, D. M. B., Luxburg, U. v., & Franz, V. H. (2021). Group decisions based on confidence weighted majority voting. *Cognitive Research: Principles and Implications*, *6*(1), 18. https://doi.org/10.1186/s41235-021-00279-0

Moré, J. J., & Sorensen, D. C. (1983). Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, *4*(3), 553–572. https://doi.org/10.1137/0904038

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817. https://doi.org/10.1287/mnsc.2021.4294

Nti, I. K., Adekoya, A. F., Weyori, B. A., & Nyarko-Boateng, O. (2022). Applications of artificial intelligence in engineering and manufacturing: A systematic review. *Journal of Intelligent Manufacturing*, *33*(6), 1581–1601. https://doi.org/10.1007/s10845-021-01771-6

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

Oyebode, O., Fowles, J., Steeves, D., & Orji, R. (2023). Machine learning techniques in adaptive and personalized systems for health and wellness. *International Journal of Human–Computer Interaction*, *39*(9), 1938–1962. https://doi.org/10.1080/10447318.2022.2089085

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans: A Publication of the IEEE Systems, Man, and Cybernetics Society*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Raina, A., McComb, C., & Cagan, J. (2019). Learning to design from humans: Imitating human designers through deep learning. *Journal of Mechanical Design*, *141*(11), 111102. https://doi.org/10.1115/1.4044256

Reason, J. (2000). Human error: Models and management. *BMJ (Clinical Research ed.)*, *320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

Regenwetter, L., Nobari, A. H., & Ahmed, F. (2022). Deep generative models in engineering design: A review. *Journal of Mechanical Design*, *144*(7), 071704. https://doi.org/10.1115/1.4053859

Rispens, S., & Demerouti, E. (2016). Conflict at work, negative emotions, and performance: A diary study. *Negotiation and Conflict Management Research*, *9*(2), 103–119. https://doi.org/10.1111/ncmr.12069

Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, *10*(3), 241–278. https://doi.org/10.1080/026999396380240

Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering Structures*, *171*(September), 170–189. https://doi.org/10.1016/j.engstruct.2018.05.084

Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), 103174. https://doi.org/10.1016/j.im.2019.103174

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Sofge, D., Lawless, W. F., & Mittu, R. (2019). AI bookie: Will a self-authorizing AI-based system take control from a human operator? *AI Magazine*, *40*(3), 79–84. https://doi.org/10.1609/aimag.v40i3.5196

Su, J., Vargas, D. V., & Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, *23*(5), 828–841. https://doi.org/10.1109/TEVC.2019.2890858

Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group & Organization Management*, *34*(2), 170–205. https://doi.org/10.1177/1059601108331218

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Villa, S., Kosch, T., Grelka, F., Schmidt, A., & Welsch, R. (2023). The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior*, *146*(September), 107787. https://doi.org/10.1016/j.chb.2023.107787

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2014). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Motivation Science*, *1*(S), 19–42. https://doi.org/10.1037/2333-8113.1.S.19

Williams, P. G., Rau, H. K., Suchy, Y., Thorgusen, S. R., & Smith, T. W. (2017). On the validity of self-report assessment of cognitive abilities: Attentional control scale associations with cognitive performance, emotional adjustment, and personality. *Psychological Assessment*, *29*(5), 519–530. https://doi.org/10.1037/pas0000361

Wilson, J., & Daugherty, P. (2018, July 1). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

Zhang, G., Chong, L., Kotovsky, K., & Cagan, J. (2023). Trust in an AI versus a human teammate: The effects of teammate identity and performance on human-AI cooperation. *Computers in Human Behavior*, *139*(February), 107536. https://doi.org/10.1016/j.chb.2022.107536

Zhang, G., Raina, A., Cagan, J., & McComb, C. (2021). A cautionary tale about the impact of AI on human design teams. *Design Studies*, *72*(January), 100990. https://doi.org/10.1016/j.destud.2021.100990

About the authors

Mo Hu is an assistant professor of Construction Science at Texas A&M-College Station. Her research mainly focuses on decision-making, human-AI interaction, and design neurocognition. She uses human subject experiments and neuroimaging techniques to explore cognition-behavior links in complex environments.

Guanglu Zhang is a research scientist in the Department of Mechanical Engineering at Carnegie Mellon University. His research interests include artificial intelligence, engineering design, GPU computing, and numerical methods.

Leah Chong is a postdoctoral associate in the Department of Mechanical Engineering at Massachusetts Institute of Technology. Her research focuses on human-AI collaboration in engineering design, computational design methods, and human-centered design, effectively and ethically harnessing the strengths of humans and AI.

Jonathan Cagan is the David and Susan Coulter Head of Mechanical Engineering and George Tallman and Florence Barrett Ladd Professor at Carnegie Mellon University. Cagan's research focuses on design automation and methods, problem solving, and medical technologies, merging AI, machine learning, and optimization methods with cognitive- and neuro-science problem solving.

Kosa Goucher-Lambert is an assistant professor in the Department of Mechanical Engineering and Jacobs Institute for Design Innovation at University of California, Berkeley. Goucher-Lambert's research is in design theory, methodology, and automation. He combines computational methods with studies of human cognition and behavior to investigate problems in engineering and design.

Appendix

A. Experience weights in the confidence model for all voting and feedback scenarios

	Outvoted for	AI #1	AI #1	AI #2	AI #2	Human	Human	Not outvote	Not outvote
	Feedback (+/-)	+	-	+	-	+	-	+	-
Confidence Models	Conditions								
Confidence in AI #1	Hi - Hi	1	0	0.59	0.47	0.58	0.61	0.56	0.44
	Lo - Hi	0.88	0.19	0.48	0.42	0.34	0.36	0.4	0.42
	Lo - Lo	0.63	0.13	0.71	0.25	0.65	0.34	0.27	0.34
Confidence in AI #2	Hi - Hi	0.26	0.01	0.87	0.21	1	0.98	0.59	0.45
	Lo - Hi	0	0.1	0.96	0	0.89	0.68	0.72	0.32
	Lo - Lo	0	0.19	0.91	0.15	0.42	0.28	0.29	0.35
Self-confidence	Hi - Hi	0.9	1	0.27	0.58	0.25	0	1	0.3
	Lo - Hi	0.52	0.68	0.55	0.74	0.26	0.29	0.65	0.25
	Lo - Lo	1	1	0.56	0.66	0.19	0	1	0
		ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8

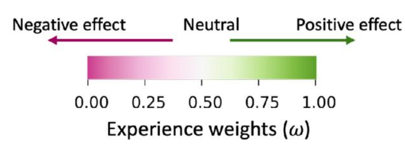

Figure A1. Experience weights for all voting and feedback scenarios.

B. Post-experiment questionnaire

(1) How good was Taylor (AI #1, left side) at designing the trusses?
-Very good -Good -Neutral -Bad -Very bad

(2) How helpful were Taylor (AI #1)'s design actions in doing this task?
-Very helpful -Helpful -Neutral -Unhelpful -Very unhelpful

(3) How good was Alex (AI #2, right side) at designing the trusses?
-Very good -Good -Neutral -Bad -Very bad

(4) How helpful were Alex (AI #2)'s design actions in doing this task?
-Very helpful -Helpful -Neutral -Unhelpful -Very unhelpful

(5) Have you been outvoted in the experiment? For instance, you voted for your design action, but both AIs voted for Taylor's or Alex's design action, or you voted for Taylor's or Alex's design action but both AIs voted for your design action.

-Yes -No

(6) If you were outvoted, and your team selected AI's (Talyor's or Alex's) action as the final design action, what was your emotional state when the feedback for the final design was advantageous/disadvantageous?

6.1. What was your emotional state in the pleasure dimension (1 = sad, 5 = neutral, 9 = happy)?

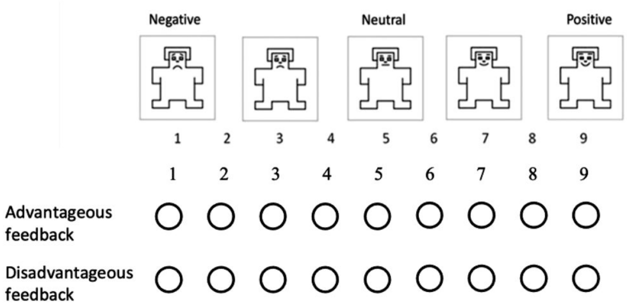

Figure A2. Emotional state in the pleasure dimension.

6.2. What was your emotional state in the arousal dimension (1 = calm, 5 = neutral, 9 = excited)?

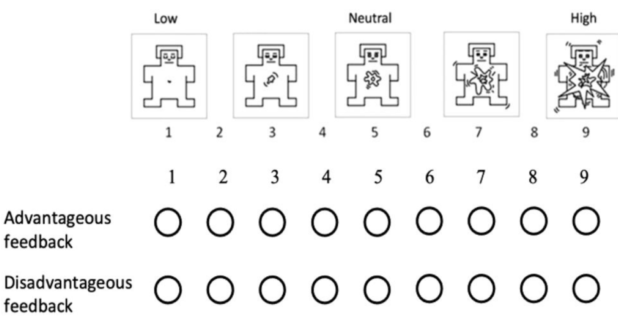

Figure A3. Emotional state in the arousal dimension.

6.3. What was your emotional state in the dominance dimension (1 = submissive, 5 = neutral, 9 = controlled)?

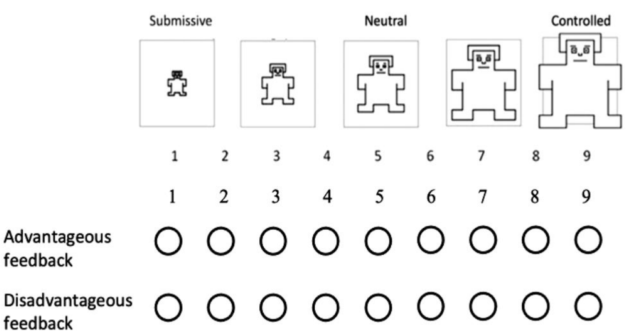

Figure A4. Emotional state in the dominance dimension.

(7) If you were outvoted, and your team selected AI's (Talyor's or Alex's) action as the final design action, what was your emotional state when the feedback for the final design was advantageous/disadvantageous?

7.1. What was your emotional state in the pleasure dimension (1 = sad, 5 = neutral, 9 = happy)?

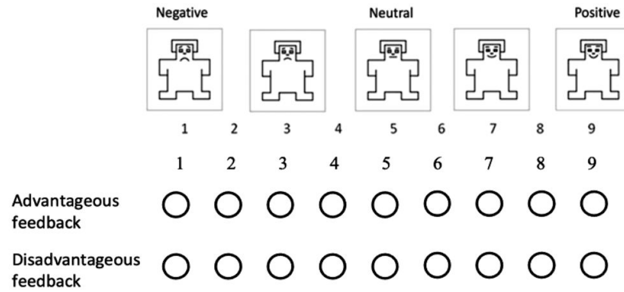

Figure A5. Emotional state in the pleasure dimension.

7.2. What was your emotional state in the arousal dimension (1 = calm, 5 = neutral, 9 = excited)?

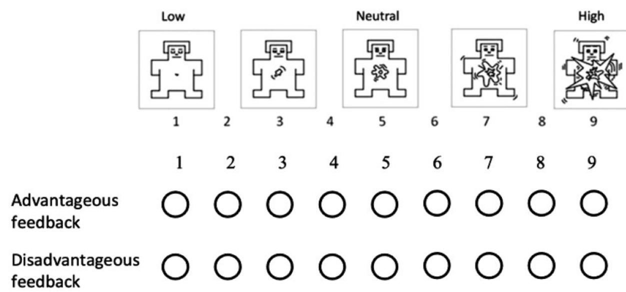

Figure A6. Emotional state in the arousal dimension.

7.3. What was your emotional state in the dominance dimension (1 = submissive, 5 = neutral, 9 = controlled)?

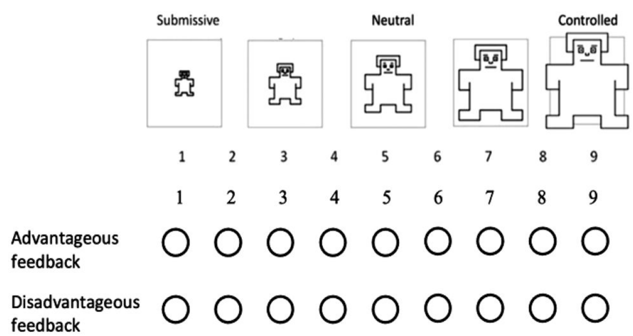

Figure A7. Emotional state in the dominance dimension.