Group Decision-Making in Robot Teleoperation: Two Heads are Better Than One

Duc-An Nguyen *
Oxford Robotics Institute
University of Oxford
Oxford, United Kingdom
annguyen@robots.ox.ac.uk

Raunak Bhattacharyya *
Yardi School of Artificial Intelligence
Indian Institute of Technology Delhi
New Delhi, India
raunakbh@iitd.ac.in

Clara Colombatto
Department of Psychology
University of Waterloo
Ontario, Canada
clara.colombatto@uwaterloo.ca

Steve Fleming
Department of Experimental Psychology
University College London
London, United Kingdom
stephen.fleming@ucl.ac.uk

Ingmar Posner
Oxford Robotics Institute
University of Oxford
Oxford, United Kingdom
ingmar@robots.ox.ac.uk

Nick Hawes
Oxford Robotics Institute
University of Oxford
Oxford, United Kingdom
nickh@robots.ox.ac.uk

***Abstract*—Operators working with robots in safety-critical domains have to make decisions under uncertainty, which remains a challenging problem for a single human operator. An open question is whether two human operators can make better decisions jointly, as compared to a single operator alone. While prior work has shown that two heads are better than one, such studies have been mostly limited to static and passive tasks. We investigate joint decision-making in a dynamic task involving humans teleoperating robots. We conduct a human-subject experiment with $N = 100$ participants where each participant performed a navigation task with two mobiles robots in simulation. We find that joint decision-making through confidence sharing improves dyad performance beyond the better-performing individual ($p < 0.0001$). Further, we find that the extent of this benefit is regulated both by the skill level of each individual, as well as how well-calibrated their confidence estimates are. Finally, we present findings on characterising the human-human dyad's confidence calibration based on the individuals constituting the dyad. Our findings demonstrate for the first time that two heads are better than one, even on a spatiotemporal task which includes active operator control of robots.**

***Index Terms*—Joint Decision-Making, Human-Robot Interaction, Teleoperation.**

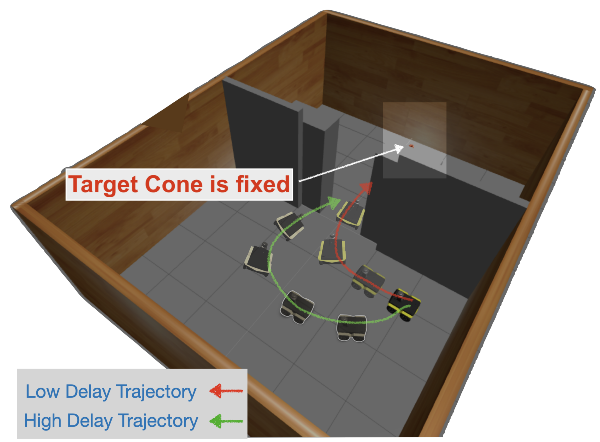


Fig. 1: Web based robot navigation simulator. The red and green curves represent trajectories driven by a participant from our user study who was experiencing low and high delay in control of the mobile robot.

I. INTRODUCTION

Human operators are increasingly collaborating with robots via teleoperation in domains such as inspection [32, 10, 15, 16, 18, 69], nuclear decommissioning [55, 17], and search and rescue [13, 21, 46, 54]. In these complex environments, operators are often faced with the decision of choosing which robot or robot controller to operate. For instance, an operator may need to select a robot from a fleet to assist in case of failure [39] or choose which robot to teleoperate [43, 36, 57] and the level of autonomy for it to operate at [56, 51, 57, 58] during missions under time pressure and uncertainty. Decisions like these are complicated by communication latencies, incomplete information, and the dynamic nature of the environment, making the choice of which robot to assist or teleoperate a challenging and cognitively demanding task. In such scenarios, single operators can become overwhelmed, especially when required to make decisions under stress [14, 47, 62, 63, 58]. Individual human operators have limitations such as cognitive overload and biases. This raises an important question: **Could two decision-makers working together make better decisions than one individual acting alone?** Collaborative decision-making may provide a solution to these limitations. While each operator has their own weaknesses, humans often possess complementary skills that, when combined, can offset each other's shortcomings. Several studies in human-human interaction (HHI) shows that collaboration can lead to better decision outcomes [31, 41, 49, 59, 64], and there is a strong case for transferring HHI findings into human-robot interaction (HRI) [74, 48, 35, 67, 45].

*Both authors contributed equally to this paper.

In the aforementioned domains, many teleoperation scenarios already involve multiple humans operating a single robot [38, 19, 12, 30]. The ratio of people to robots directly affects the human-robot interaction in such systems. For example, in the mobile robot search and rescue operations using telerobots, the operator-to-robot ratio is commonly 2 to 1 or higher [53, 68]. In control rooms for Unmanned Aerial Vehicle missions, several operators are needed to operate a single drone. In these scenarios, specific roles are often assigned to different team members, such as a "pilot" responsible for navigation and control, and a sensor/payload operator managing cameras and other equipment [22]. In the multi-UAV control setting, Hughes et al. showed that combining multiple human operators with partially autonomous robots can create more robust and effective systems than either working alone. In these multi-operator, single-robot scenarios, team dynamics [11, 78, 33], communication [34, 73] and coordination strategies [2, 23] are all critical to successful team performance.

In HHI collaborative cognition, a line of research on human-human dyadic joint decision-making has shown that teams can outperform individuals under specific conditions, particularly when the joint decisions are guided by the Maximum Confidence Slating (MCS) approach [4, 5, 6, 20, 24]. This approach prioritizes the decision in which the participant has the *higher confidence*, assuming that each individual can monitor their own performance and can communicate their confidence accurately [3, 40]. While MCS has been shown to improve performance in tasks such as visual perception or knowledge-based tasks, its application to active, dynamic decision-making in robotic control tasks has not been adequately studied. We address this gap by investigating the performance of MCS in a spatio-temporal task, where human operators control robots in a simulated environment. Unlike previous research, which focused on passive information processing, our study involves humans actively gathering and interpreting information to make real-time decisions.

Specifically, in our user study, we conducted experiments with 100 participants, each controlling two different robots in an online simulation. The environments and control latencies of the robots were altered, with one robot exhibiting a more favorable delay. Participants had to choose which robot was better for the task, based on the robot's responsiveness and their own confidence in the choice. We then applied the MCS approach to compare the accuracy of joint decision-making of dyads being virtually paired with individual performance, using an *accuracy gain*[1] metric.

To the best of our knowledge, this is the first study to apply MCS in a scenario where humans must actively control robots rather than passively receiving information. Our findings reveal the following key insights into the effectiveness of the MCS-based decision-making process in joint decision-making within robotics teleoperation domains:

1) **Impact of Confidence on Accuracy:** Joint decisions determined by high confidence via MCS were more accurate than those made by the highest-performing individual in the dyad. Conversely, joint decisions based on low-confidence inputs were less reliable than random choices, emphasizing the need to avoid low-confidence responses.
2) **Effect of Performance Discrepancy:** This is the first time we have observed that larger performance gaps between dyad members result in smaller accuracy gains from MCS. Pairing participants with similar performance levels yielded significantly higher accuracy, highlighting that MCS is more effective when participants have comparable competence.
3) **Influence of Confidence Calibration:** Confidence calibration [25] significantly impacted MCS accuracy gains. For individuals with above-average confidence calibration, MCS-based accuracy gains were stable regardless of calibration differences. For those with below-average calibration, pairing similarly calibrated individuals led to negative accuracy gains, while teaming individuals with diverse calibration levels improved the accuracy of the dyad. This marks the first time such analyses have been conducted on joint decision-making in dynamic tasks.
4) **Correlation with Dyadic Confidence Calibration:** Accuracy gains from MCS were positively correlated with dyadic confidence calibration. Higher overall confidence calibration of the dyad led to better decision accuracy, further underscoring the importance of confidence calibration in joint decision-making.

II. RELATED WORK

A. Multi-operator Teleoperation

Multiple Operator Single Robot (MOSR) systems have emerged as a promising approach, particularly in the field of semiautonomous teleoperation [70, 71, 77, 72, 76, 75]. Reed and Peshkin highlighted the potential benefits of MOSR systems via the studies of haptic human-human interaction in joint object manipulation tasks. It was shown that task performance of two humans solving a haptic task collaboratively is higher than that of a single operator performing the same task. The authors suggests that adding an additional human operator to a classical teleoperation scheme could have a positive effect on task performance [52].

Another example is the distributed teleoperation system developed by Goldberg et al., where multiple users, each at a different location, simultaneously controlled an industrial robot arm over the internet. Their client-server system enabled multiple users to share control of a single robot, demonstrating how collaboration among users can enhance performance. Importantly, this research showed that human groups outperformed individuals when facing noisy environments in MOSR systems [26].

[1]Accuracy gain is defined as the accuracy increase/decrease difference between the accuracy (%) of decision made by the MCS-based joint decision-making agent as compared with the higher performing individual in the dyad

(a) (b) (c)

Fig. 2: Screenshots showing driving trials scenario and question queries (2a: environment setup; 2b: robot choices, 2c: user confidence levels) for each participant. There are total of 100 trials per participant.

B. Joint Decision-Making in Human Dyads

Joint decision-making within human dyads has been extensively studied in the context of visual perception [3, 4, 5, 7, 61, 20]. Bahrami et al. conducted experiments where participants, working in dyads, engaged in a two-alternative forced-choice task, deciding which of two briefly presented stimuli contained an oddball target. Participants initially made individual decisions before sharing their conclusions. In cases of disagreement, they discussed the matter until reaching a joint decision. The results confirmed that, given equal visual sensitivity, two individuals were indeed better than one, provided they could communicate freely. The authors suggested that this "two-heads-are-better-than-one" (2HABT1) effect depends on each participant's ability to monitor their own performance accuracy and accurately communicate their confidence level.

Further studies by Koriat demonstrated that the advantage of joint decision-making remains even when participants are unable to communicate directly [6, 50, 65]. It is shown that decisions guided by the confidence heuristic—where the most confident choice is selected—can be just as accurate as decisions reached through direct interaction, especially for individuals with comparable reliability. This highlights the potential of confidence as a powerful tool in improving joint decision-making within dyads. Over time, the accuracy of collective decisions becomes indistinguishable between conditions with and without feedback. Although the learning process is slower without feedback, the eventual collective benefit matches that of situations where feedback is provided [6]. In line with these discoveries, studies in human-human joint decision-making on varied tasks such as threat detection by observing video feeds [9], detecting fake news [28, 60] deciding rank ordering between items on a survival situation task [29], and breast and skin cancer diagnosis [42] have also shown that higher confidence decision selection leads to higher accuracy.

Building on the body of related work described above, we identify a gap in the research regarding dyadic joint decision-making in teleoperated robotic tasks, and formulate the following key research questions to guide our investigation into how humans can make better joint decisions through MCS in teleoperational robotics:

RQ1: How does individual confidence affect the accuracy of joint decisions in dyadic settings, and what are the benefits of joint decision-making using MCS?

RQ2: How do performance discrepancies between dyad members affect the accuracy gains of MCS-based joint decisions?

RQ3: How does the confidence calibration of participants influence the accuracy gains of MCS-based joint decisions?

RQ4: What is the relationship between dyadic confidence calibration and the accuracy of dyadic MCS-based joint decisions?

III. METHODOLOGY

We conducted an online study, approved by the University of Oxford Research Ethics Committee, to investigate how humans make joint decisions when selecting robot controllers.

A. Experiment Setup

The experiment tasked participants with navigating two visually identical simulated robots through a narrow gap using an interface shown in Figure 1. The environment consisted of 24 different conditions, combining 6 possible initial robot poses with 4 doorway configurations. A fixed goal location was marked by a traffic cone, with the initial robot pose and door configuration randomly sampled for each trial. Participants were recruited via the Prolific Platform, which provided an overview of the study and ethics approval information. Those who agreed to participate and were at least 18 years old were shown a detailed video explaining the interface components and demonstrating example trials. Before the main experiment, participants completed 5 practice trials to familiarize themselves with the keyboard controls.

B. Procedure

The main experiment consisted of 100 trials. In each trial, participants controlled each robot for 6 seconds in a random order. This duration was determined through a separate pilot study as the minimum time needed for decision-making. To create a perceptible difference between the robots, we injected time delays into their controls. Randomly, one was assigned a fixed 50 ms delay, and the other a variable delay (70–150 ms). We implemented a two-down one-up staircase[2] procedure to adjust task difficulty. This procedure ensured that the task was not so easy as to enable ceiling performance and not so hard to make participant performance close to random [44]. This precluded participants from being overly confident or underconfident on the task. During each trial, participants controlled each robot sequentially, with a brief pause, and selected the robot with the lower delay, rating their confidence on a four-point Likert scale (1:lowest, 4:highest). This experimental design allowed us to assess both decision-making accuracy and confidence calibration in human-robot interaction tasks under varying degrees of difficulty. Figure 2 depicts the trial procedure.

The data were later used to pair individuals into dyads virtually. For example, Participant 1 encountered 20 trials with the delay-pair condition: 50 ms delay for Robot 1 and 70 ms delay for Robot 2. Participant 2 encountered 15 trials with the same delay condition. Therefore, we could pair up to a total of $\frac{20\times15}{2} = 150$ trials, where Participants 1 and 2 were able to make virtually joint decisions.

C. Data Collection

To ensure data quality, we implemented exclusion criteria based on task performance and engagement. Participants with accuracy below 65% or those who gave the same confidence rating for over 95 out of 100 trials were excluded. This threshold helped retain engaged participants while excluding those who were inattentive or ineffective. The exclusion criteria and performance metrics are consistent with standard procedures in human self-confidence research [66, 8, 1], ensuring our results are comparable with existing literature in the field. After applying these criteria, our final analysis included 80 participants (47 males, 33 females), with a mean age of 37 years (SD = 11) and an average experiment duration of 1.2 hours (SD = 0.25). Participants were compensated an average of USD 7.6.

D. Metrics

We quantified task performance using two primary metrics:

1) Accuracy: The proportion (%) of correct selections of the lower-delay robot.
2) Confidence calibration: AUROC2 computation measures how well participants' confidence ratings align with their actual performance. Algorithm 1 presents a detailed computation of AUROC2. [25]

These metrics allowed us to evaluate both the participants' ability to discriminate between the robots based on delay and their metacognitive awareness of their performance. By combining accuracy and confidence calibration measures, we aimed to gain comprehensive insights into human decision-making and self-assessment in this human-robot interaction task. This enables us to understand not just how well participants performed, but also how accurately they judged their own performance across varying levels of task difficulty.

Algorithm 1: Confidence Calibration (AUROC2)

Input: $correct$: vector of size $1 \times n_{trials}$, with 0 for error and 1 for correct trials
$conf$: vector of size $1 \times n_{trials}$, with confidence ratings from 1 to $N_{ratings}$
$N_{ratings}$: number of available confidence levels
Output: $auroc2$: type-2 area under the ROC curve
Initialize $i \leftarrow N_{ratings} + 1$
for $c \leftarrow 1$ **to** $N_{ratings}$ **do**
  $H2[i-1] \leftarrow \text{count}(conf = c \wedge correct) + 0.5$
  $FA2[i-1] \leftarrow \text{count}(conf = c \wedge \neg correct) + 0.5$
  $i \leftarrow i - 1$
end
Normalize $H2 \leftarrow H2/\sum(H2)$
Normalize $FA2 \leftarrow FA2/\sum(FA2)$
Compute cumulative sums:
 $csum\_H2 \leftarrow [0, \text{cumsum}(H2)]$,
 $csum\_FA2 \leftarrow [0, \text{cumsum}(FA2)]$
Initialize $i \leftarrow 1$
for $c \leftarrow 1$ **to** $N_{ratings}$ **do**
  $k[i] \leftarrow (csum\_H2[c+1] - csum\_FA2[c])^2 - (csum\_H2[c] - csum\_FA2[c+1])^2$
  $i \leftarrow i + 1$
end
Compute $auroc2 \leftarrow 0.5 + 0.25 \times \sum(k)$
return $auroc2$

IV. RESULTS

From the data obtained from 80 participants, we combine pairwise participant data to form virtual dyads [40]. We obtained 3160 virtual dyads $\binom{80}{2}$. For each virtual dyad, we selected those trials where both the participants faced the same delay pair. From these, we eliminated those trials where both participants agreed on their choice of robot, and further analysed those trials where the participants differed in their choice such that a joint decision was required.

The pairing process (III-B) was repeated for all delay-pair conditions and 4,148 virtual joint decision-making trials were generated from real data. Of these, 1,932 trials involved disagreements, requiring joint decisions for this dyad.

Within each dyad, the member with a higher percentage of correct responses was designated as high-performing (HP), while the other was designated as low-performing (LP). Additionally, three dummy participants were created: dummy high-confidence (D-HC), dummy low-confidence (D-LC), and dummy random (D-Random). For each trial, the response of the participant who indicated higher confidence was selected as D-HC participant, and the other as the D-LC participant. **D-HC corresponds to the Maximum Confidence Slating (MCS) approach of joint decision-making**. D-Random represented a joint decision-maker who picks randomly between

[2] The two-down one-up staircase is a procedure where the task was made harder after 2 consecutive successful choices by decreasing the delay by one level step, and the task was made easier after 1 failure by increasing the delay by one level step.

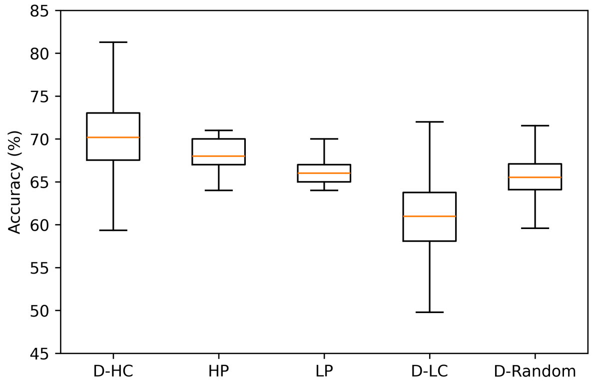


Fig. 3: Accuracy in robot selection for high-confidence, higher-performing, low-confidence, lower-performing, and random participants from the paired dyads.

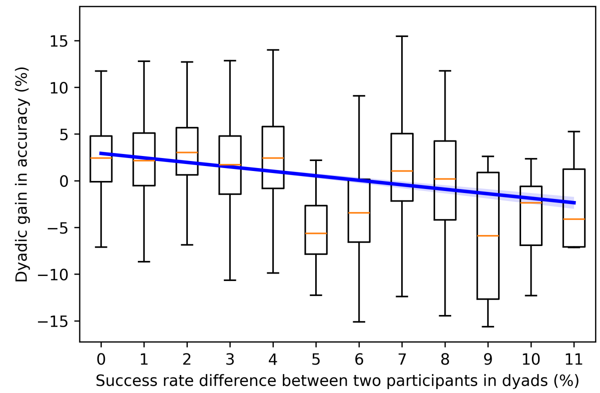


Fig. 4: Variation in accuracy gain as participants of different success rates are paired together. Pairing similar performing participants leads to higher benefit.

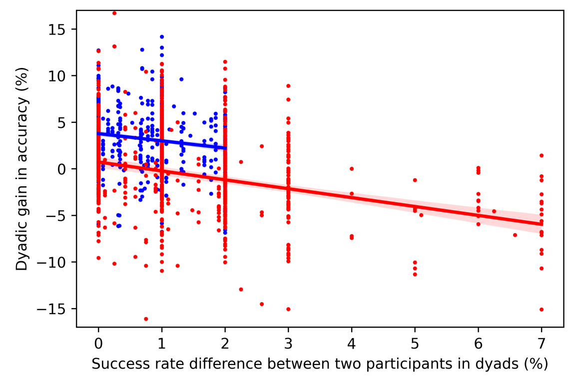


Fig. 5: Variation in accuracy gain as participants of different success rates are paired together. Blue: both participants being below the mean skill level, and red: both participants being above the mean skill level.

the choices made by both the participants in the dyad. Accuracy was then calculated for these five participants.

We then present our analysis formulated from the four backbone research questions. Formally, the independent variables are (1) individual confidence, (2) performance discrepancy, and (3) confidence calibration. The dependent variables include joint decision accuracy, accuracy gains from MCS, and the dyadic AUROC2 value measuring the dyad's confidence calibration.

A. Impact of Confidence on Accuracy

RQ1: *How does individual confidence affect the accuracy of joint decisions in dyadic settings, and what are the benefits of joint decision-making using MCS?*

Figure 3 illustrates the accuracy of robot selection across all five participant types, based on 3,160 virtual dyads. We assessed the performance of D-HC against other participants using the two-sample t-test. The results show that D-HC exhibited significantly higher accuracy compared to HP (Highest Performing individual), $(t(3158) = 8.36, p < 0.0001)$. This confirms that decisions made using the Maximum Confidence Slating (MCS) approach provide a significant advantage over those made by the highest-performing individual within the dyad. The accuracy of D-HC was also significantly greater than that of D-Random, $(t(3158) = 33.48, p < 0.0001)$. Interestingly, D-Random performed better than D-LC, $(t(3158) = 33.48, p < 0.0001)$. This suggests that, in any given scenario, a low-confidence response should not be favored. Additionally, we confirmed that D-HC had significantly higher accuracy than D-LC, $(t(3158) = 62.06, p < 0.0001)$, reinforcing that high-confidence decisions lead to better outcomes than low-confidence ones.

Through this analysis, we validated **H1: MCS is effective in robot teleoperation control selection task. Picking the higher confidence decisions within a dyad results in more accurate joint decisions than those made by the highest-performing individual alone. Conversely, low-confidence decisions yield accuracy levels comparable to random choices.**

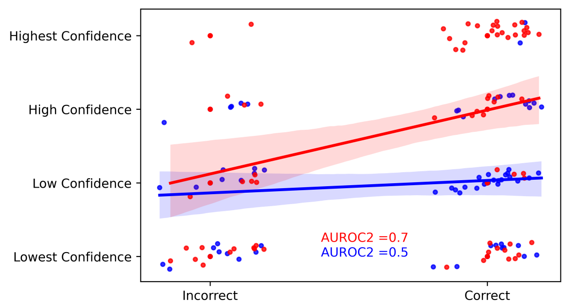
B. Effect of Performance Discrepancy

RQ2: *How do performance discrepancies between dyad members affect the accuracy gains from MCS-based joint decisions?*

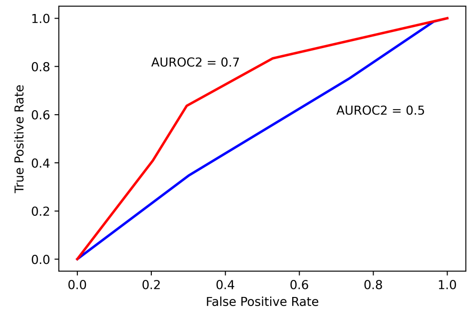
While MCS-based joint decision-making delivers a benefit overall, specific ways of pairing participants provides further insights. In this section, we analyse the impact of pairing participants according to their performance level. We sorted the 3160 virtual dyads according to the **difference** in the performance of both individuals in each dyad, as measured using two factors:

1) Success rate $(\%)$ over the 100 trials.
2) Settling task difficulty level via the two-down one-up procedure.

For every virtual dyad, we extracted the corresponding *accuracy gain* [6]. Figure 4 shows the impact of difference in performance level on the resulting accuracy gain of the dyad. The difference in success percentage between the individuals in each virtual dyad ranged from 0 to 11%. As the dissimilarity in success rate increases, the resulting accuracy gain decreases (regression coefficient : $r = -0.48, p < 0.0001$).



(a) Correlation between correctness and confidence.



(b) Corresponding AUROC2 curves.

Fig. 6: Confidence calibration of two different participants. 6a: Plot of trial result vs. confidence rating. The line is the regression fit with 95% confidence interval. 6b: ROC curves for both participants. The participant represented by red has better calibrated confidence as compared to the participant represented by blue.

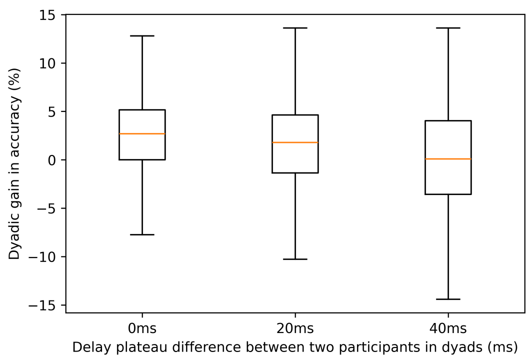


Fig. 7: Variation in accuracy gain as participants of different performance levels are paired together. Performance level is assessed by the difficulty of the task as established by delay plateau difference. Pairing similarly performing participants leads to higher accuracy gain.

The accuracy gain corresponding to the dyads with no difference in success rate (i.e., identical skill level) was significantly higher than both dyads with a 5% difference in success rate ($t(1008) = 7.1, p < 0.0001$)), as well as dyads with a 10% difference in success rate (($t(694) = 8.35, p < 0.0001$)). Pairing participants with a success rate difference of more than 8%, leads to a negative accuracy gain, i.e., MCS causes poorer joint performance as compared to the better performing individual.

For a more granular understanding of the impact of skill difference (similarly skilled can correspond to both being highly skilled or both being poorly skilled), we divided the participant data into two groups: above mean skill level (success percentage $> 67\%$), and below mean skill level (success percentage $< 67\%$). Figure 5 shows the variation in accuracy gain with difference in performance of the two individuals for the two groups of participants. While the trend is similar for both groups (blue: both participants being below the mean, and red: both participants being above the mean), the absolute values of the accuracy gains show that when both participants are below mean performance the benefit provided by MCS-based joint decision is higher. When both participants are above mean performance, the benefit is more observable for participants of similar performance level (lower success rate difference).

Figure 7 shows the impact of pairing based on performance, measured by the task's settled difficulty level. Virtual dyads are grouped into three categories based on the delay difference faced by the two participants when task difficulty plateaus. A 0 ms value on the x-axis indicates similar performance between participants, while a 40 ms value indicates differing performance levels. We see that the resulting accuracy gain is again higher for participants with similar performance level. The accuracy gain is higher for the virtual dyad with similarly performing participants (0 ms) as compared to both 20 ms ($t(5752) = 9.67, p < 0.0001$) as well as 40 ms ($t(3280) = 11.04, p < 0.0001$).

This analysis supports **H2: Similar task performance leads to higher accuracy gains. Specifically, the accuracy gains from MCS are greater in dyads where the performance levels of the members are similar, compared to dyads with significant performance discrepancies. This is the first time it is observed that larger performance gaps between dyad members result in smaller accuracy gains.**

C. Influence of Confidence Calibration

RQ3: *How does the confidence calibration of participants influence the accuracy gains from MCS-based decisions?* To find the answer to this, we paired participants according to confidence calibration and analysed the confidence calibration of participants to gain further insights into the performance of the MCS approach to joint decision-making.

Figure 6 illustrates the variation of confidence rating with the correctness of response for two different participants at either end of the spectrum of AUROC2 values. The (red) participant has better calibration (AUROC2 = 0.7) than the (blue) participant (AUROC2 = 0.5). Figure 6a shows a plot of trial result versus associated confidence value and Figure 6b shows the corresponding ROC curves for both the participants.

To analyze the impact of confidence calibration similarity on accuracy gain, participants were grouped based on their AUROC2 values: above the mean (AUROC2 > 0.6) and below

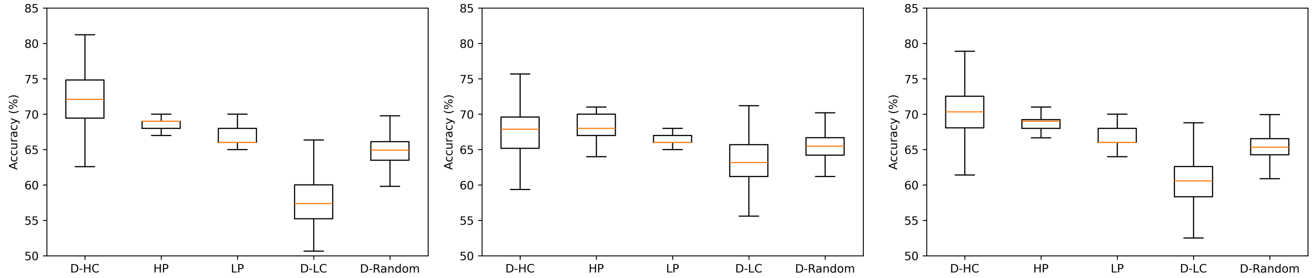


Fig. 8: Choice accuracy (%) of different combinations of participants according to their confidence calibration: 8a: both well calibrated (288 pairs), 8b: both poorly calibrated (242 pairs), and 8c: well calibrated with poorly calibrated (264 pairs).

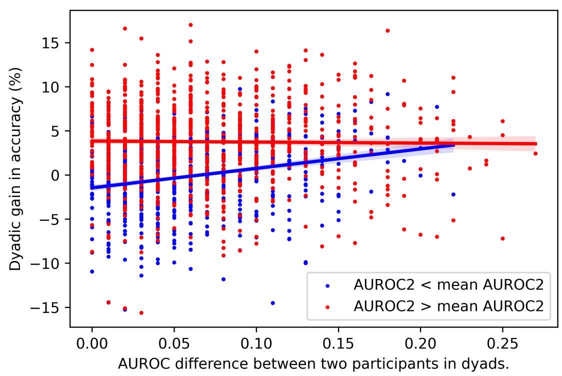


Fig. 9: Variation in accuracy gain as participants of different confidence calibration are paired together. For well-calibrated individuals (red), pairing participants of similar confidence calibration leads to higher benefit. Pairing similarly calibrated participants who are poorly calibrated (blue) does not yield a benefit.

the mean (AUROC2 < 0.6).We then paired participants within the same group and analysed the impact on accuracy gain. Figure 9 shows that for participants with AUROC2 above the mean, MCS-based joint decisions with similarly calibrated pairs improved accuracy. However, for those below the mean, pairing similar participants led to negative accuracy gain, with MCS performing worse than the better individual.

We further investigated the impact of pairing participants according to confidence calibration by selecting the top participants from the group with AUROC2 values above the mean (AUROC2 > 0.65: well-calibrated, 25 participants), and bottom participants from the group with AUROC2 values below the mean (AUROC2 < 0.55: poorly-calibrated, 26 participants). We formed virtual dyads by pairing participants according to their confidence calibration: both well-calibrated participants (288 virtual dyads); both poorly-calibrated participants (242 virtual dyads); and well-calibrated participant paired with poorly-calibrated participant (264 virtual dyads).

Figure 8 presents the performance accuracy for the five types of participants (HP, LP, D-HC, D-LC, and D-Random) in the virtual dyads. Figure 8a shows the accuracy when both individuals in the dyad are well-calibrated in terms of confidence. The performance of D-HC was significantly better than HP, $(t(574) = 1.39, p < 0.0001)$. This demonstrates that pairing two well-calibrated individuals results in an increased choice's accuracy compared to the higher performance individual. However, this is not the case when both individuals are poorly calibrated. Figure 8b illustrates the accuracy for dyads with two poorly calibrated individuals, where HP outperformed D-HC, $(t(482) = 6.08, p < 0.0001)$. This suggests that in such cases, it is preferable to rely on the more performant individual's choice. Figure 8c presents the results when a well-calibrated participant is paired with a poorly calibrated one. In this case, D-HC outperformed HP, $(t(526) = 10.86, p < 0.0001)$, indicating that the joint decision was more accurate when at least one individual had good confidence calibration.

Through this analysis, we confirmed **H3: Participants with above-average confidence calibration achieve stable accuracy gains from MCS, regardless of the calibration differences between dyad members. Conversely, participants with below-average calibration show improved accuracy when paired with individuals who have diverse calibration levels.**

D. Dyadic Confidence-Calibration Correlation

RQ4: *What is the relationship between dyadic confidence calibration and the accuracy of MCS-based joint decisions?*

We further investigated the behavior of resulting dyads formed from pairing individuals together. We computed dyadic confidence calibration using Koriat's method [40], applying the AUROC2 computation (Algorithm. 1) to the concatenated decision-confidence pairs of both participants. For instance, Participant 1 completed 20 trials and Participant 2 completed 15 trials under the same delay condition (50 ms for Robot 1, 70 ms for Robot 2). Of the combined 35 trials, 6 with identical robot choices were excluded, leaving 29 trials for dyadic confidence calibration computation. Figure 10 shows the trial result vs. confidence rating for an example participant dyad, along with the associated AUROC2 value.

The calibration of individuals in a dyad affects the resulting dyad's confidence calibration. Figure 11 shows dyadic confidence calibration based on individual calibrations. Statistical analysis reveals that dyads of two well-calibrated individuals

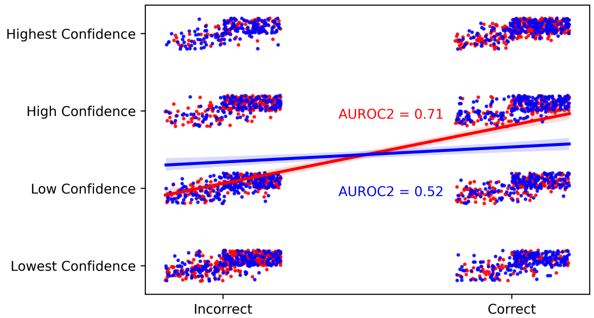


Fig. 10: Trial results versus confidence ratings for example virtual dyads. Red: a dyad formed by pairing well-calibrated participants. Blue: a dyad formed by pairing poorly calibrated participants.

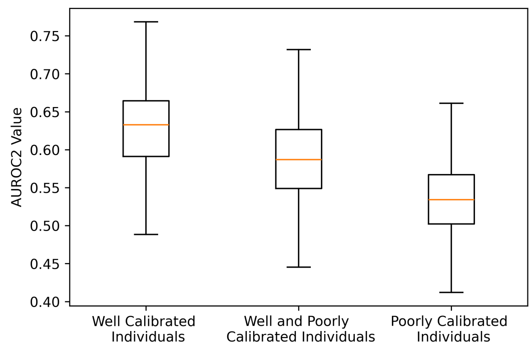


Fig. 11: Variation in dyadic confidence calibration based on how participants are paired according to their individual confidence levels.

have higher AUROC2 than those with both poorly calibrated individuals $(t(1012) = 28.74, p < 0.001)$ or mixed-calibration dyads $(t(1078) = 13.28, p < 0.001)$.

Finally, our findings highlight the significant impact of dyadic AUROC2 on joint decision-making performance. Figure 12 shows a strong correlation between accuracy gain and dyadic confidence calibration, with improved calibration leading to higher joint decision accuracy. This underscored the importance of well-calibrated confidence within the dyad for the effectiveness of the Maximum Confidence Slating (MCS) approach.

This analysis supports our fourth hypothesis **H4: Dyads with higher overall confidence calibration demonstrate better decision accuracy when using MCS-based joint decision, compared to dyads with lower calibration. This finding reinforces the value of pairing individuals with well-calibrated confidence for optimal team performance.**

V. DISCUSSION

Summary:

In this research, we investigated human-human dyad joint decision-making in a robot teleoperation task, focusing on how maximum confidence slating (MCS) choice selection impacts decision accuracy. To our knowledge, this is the first study applying MCS-based joint decision-making in a dynamic, spatiotemporal task involving active robot control.

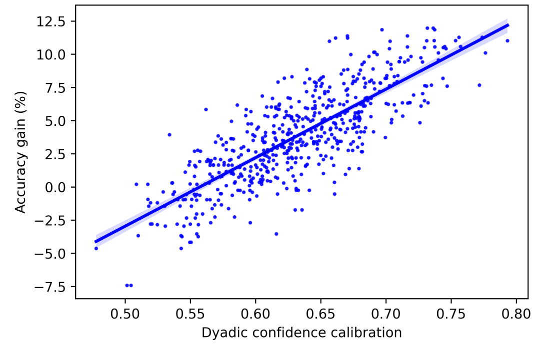


Fig. 12: Dyadic confidence calibration versus accuracy gain. Higher dyadic confidence calibration leads to an increased accuracy gain.

Our results showed that the accuracy of dyad joint decisions was significantly higher than that of the more skilled individual in the pair. Our findings emphasise the importance of skill similarity and confidence calibration in achieving better outcomes for human collaboration in robotic tasks, establishing a foundational understanding of MCS's role in the human-robot interaction domain. The effectiveness of MCS in this context highlights its potential for improving joint decision-making in human-IDS (Intelligent Decision Support) system dyads. This initial work also reveals the potential of MCS for real-world applications requiring critical, time-sensitive decisions. By leveraging confidence as a low-cost metric, MCS combines two operators' confidence levels to improve task efficiency and decision accuracy.

Limitations and Research Directions:

The scope of our study was constrained to a specific robot teleoperation and controller selection task, where the primary decision involved selecting between two robots with different control delays. Our paper looked into virtual dyads scenarios, and in real joint decision-making between two people social effects like the perceived competence of oneself and of the other, differences in status can play a big role. Future research can expand to encompass a broader range of tasks and decision-making scenarios.

One key area is extending our study to human-AI dyads. We plan to develop an Intelligent Decision Support (IDS) and AI systems tailored to the robot teleoperation task. By comparing findings from human-AI dyads with those from human-human dyads, we can gain valuable insights into how human confidence calibration influences decision-making when interacting with AI systems in robotic scenarios. Through these research directions, we hope to contribute to the integration of AI systems and development of more effective collaborative Human-Robot Interaction (HRI) settings.

VI. ACKNOWLEDGEMENT

This work was supported by the EPSRC Programme Grant 'From Sensing to Collaboration' (EP/V000748/1). Raunak Bhattacharyya acknowledges support from the Yardi School of AI publication grant.

References

[1] Ajith Anil Meera and Pablo Lanillos. Towards metacognitive robot decision making for tool selection. In *International Workshop on Active Inference*. Springer, 2023.

[2] Reuben M Aronson and Elaine Schaertl Short. Intentional user adaptation to shared control assistance. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2024.

[3] Bahador Bahrami, Karsten Olsen, Peter E Latham, Andreas Roepstorff, Geraint Rees, and Chris D Frith. Optimally interacting minds. *Science*, 329(5995):1081–1085, 2010.

[4] Bahador Bahrami, Karsten Olsen, Dan Bang, Andreas Roepstorff, Geraint Rees, and Chris Frith. Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making. *Journal of Experimental Psychology: Human Perception and Performance*, 38(1):3–8, February 2012. ISSN 1939-1277, 0096-1523. doi: 10.1037/a0025708.

[5] Bahador Bahrami, Karsten Olsen, Dan Bang, Andreas Roepstorff, Geraint Rees, and Chris Frith. What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594):1350–1365, May 2012. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2011.0420.

[6] Dan Bang, Riccardo Fusaroli, Kristian Tylén, Karsten Olsen, Peter E. Latham, Jennifer Y.F. Lau, Andreas Roepstorff, Geraint Rees, Chris D. Frith, and Bahador Bahrami. Does interaction matter? testing whether a confidence heuristic can replace interaction in collective decision-making. *Consciousness and Cognition*, 26, 2014. doi: 10.1016/j.concog.2014.02.002.

[7] Dan Bang, Laurence Aitchison, Rani Moran, Santiago Herce Castanon, Banafsheh Rafiee, Ali Mahmoodi, Jennifer Y. F. Lau, Peter E. Latham, Bahador Bahrami, and Christopher Summerfield. Confidence matching in group decision-making. *Nature Human Behaviour*, 1(6):0117, 2017. ISSN 2397-3374. doi: 10.1038/s41562-017-0117.

[8] Raunak Bhattacharyya, Duc An Nguyen, Clara Colombatto, Stephen Fleming, Ingmar Posner, and Nick Hawes. Towards intelligent decision support systems in robotics: Investigating the role of self-confidence calibration in joint decision-making. In *AAAI Spring Symposium Series*, 2024.

[9] Saugat Bhattacharyya, Davide Valeriani, Caterina Cinel, Luca Citi, and Riccardo Poli. Target detection in video feeds with selected dyads and groups assisted by collaborative brain-computer interfaces. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 159–162. IEEE, 2019.

[10] Matthew Budd, Paul Duckworth, Nick Hawes, and Bruno Lacerda. Bayesian reinforcement learning for single-episode missions in partially unknown environments. In *Conference on Robot Learning (CoRL)*, 2023.

[11] Jennifer Burke, Matt Lineberry, Kevin S Pratt, Meng Taing, Robin Murphy, and Brian Day. Toward developing hri metrics for teams: Pilot testing in the field. *Metrics for Human-Robot Interaction*, 21, 2008.

[12] Yifan Cai, Abhinav Dahiya, Nils Wilde, and Stephen L. Smith. Scheduling operator assistance for shared autonomy in multi-robot teams. (arXiv:2209.03458), 2022. URL http://arxiv.org/abs/2209.03458.

[13] Jennifer Casper and Robin R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 33(3):367–385, 2003.

[14] Jessie YC Chen and Michael J Barnes. Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, 54(2):157–174, 2012.

[15] Manolis Chiou, Rustam Stolkin, Goda Bieksaite, Nick Hawes, Kimron L Shapiro, and Timothy S Harrison. Experimental analysis of a variable autonomy framework for controlling a remotely operating mobile robot. In *IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2016.

[16] Manolis Chiou, Nick Hawes, and Rustam Stolkin. Mixed-initiative variable autonomy for remotely operated mobile robots. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(4):1–34, 2021.

[17] Manolis Chiou, Georgios-Theofanis Epsimos, Grigoris Nikolaou, Pantelis Pappas, Giannis Petousakis, Stefan Mühl, and Rustam Stolkin. Robot-assisted nuclear disaster response: Report and insights from a field exercise. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[18] Javier Corredor, Jorge Sofrony, and Angelika Peer. Decision-making model for adaptive impedance control of teleoperation systems. *IEEE Transactions on Haptics*, 10(1):5–16, 2016.

[19] Abhinav Dahiya, Nima Akbarzadeh, Aditya Mahajan, and Stephen L. Smith. Scalable operator allocation for multi-robot assistance: A restless bandit approach. (arXiv:2111.06437), 2021. URL http://arxiv.org/abs/2111.06437.

[20] Benedetto De Martino, Stephen M Fleming, Neil Garrett, and Raymond J Dolan. Confidence in value-based choice. *Nature Neuroscience*, 16(1):105–110, 2013. doi: 10.1038/nn.3279.

[21] Lorin D Dole, David M Sirkin, Robin R Murphy, and Clifford I Nass. Robots need humans in the loop to improve the hopefulness of disaster survivors. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2015.

[22] Jill L Drury, Laurel Riek, and Nathan Rackliffe. A decomposition of uav-related situation awareness. In *ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI)*, 2006.

[23] Saad Elbeleidy, Terran Mott, and Tom Williams. Practical, ethical, and overlooked: Teleoperated socially assistive robots in the quest for autonomy. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022.

[24] Stephen M. Fleming and Nathaniel D. Daw. Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological Review*, 124(1):91–114, 2017.

[25] Stephen M Fleming and Hakwan C Lau. How to measure metacognition. *Frontiers in Human Neuroscience*, 8:443, 2014.

[26] Ken Goldberg, Billy Chen, Rory Solomon, Steve Bui, Bobak Farzin, Jacob Heitler, Derek Poon, and Gordon Smith. Collaborative teleoperation via the internet. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 2. IEEE, 2000.

[27] Ken Goldberg, Dezhen Song, Y Khor, David Pescovitz, Anthony Levandowski, J Himmelstein, Janice Shih, Annamarie Ho, Eric Paulos, and J Donath. Collaborative online teleoperation with spatial dynamic voting and a human" tele-actor". In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 2. IEEE, 2002.

[28] Douglas Guilbeault, Samuel Woolley, and Joshua Becker. Probabilistic social learning improves the public's judgments of news veracity. *Plos One*, 16(3):e0247487, 2021.

[29] Daisuke Hamada, Masataka Nakayama, and Jun Saiki. Wisdom of crowds and collective decision-making in a survival situation with complex information integration. *Cognitive Research: Principles and Implications*, 5(1): 1–15, 2020.

[30] Zhao Han, Yifei Zhu, Albert Phan, Fernando Sandoval Garza, Amia Castro, and Tom Williams. Crossing reality: Comparing physical and virtual robot deixis. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2023.

[31] Tsutomu Harada. Three heads are better than two: Comparing learning properties and performances across individuals, dyads, and triads through a computational approach. *PLOS ONE*, 16, 2021. doi: 10.1371/journal.pone.0252122.

[32] Nick Hawes, Christopher Burbridge, Ferdian Jovan, Lars Kunze, Bruno Lacerda, Lenka Mudrova, Jay Young, Jeremy Wyatt, Denise Hebesberger, Tobias Kortner, et al. The strands project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine*, 24(3):146–156, 2017.

[33] Hooman Hedayati, Michael Walker, and Daniel Szafir. Improving collocated robot teleoperation with augmented reality. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018.

[34] Pamela J Hinds, Teresa L Roberts, and Hank Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 2004.

[35] Guy Hoffman and Cynthia Breazeal. Cost-based anticipatory action selection for human–robot fluency. *IEEE Transactions on Robotics*, 23(5):952–961, 2007.

[36] Ryan Hoque, Lawrence Yunliang Chen, Satvik Sharma, Karthik Dharmarajan, Brijen Thananjeyan, Pieter Abbeel, and Ken Goldberg. Fleet-dagger: Interactive robot fleet learning with scalable human supervision. In *Conference on Robot Learning (CoRL)*, 2023.

[37] Thomas Hughes. Human-automation coordination in multi-uav control. In *AIAA Guidance, Navigation and Control Conference and Exhibit*, page 6315, 2008.

[38] Tianchen Ji, Roy Dong, and Katherine Driggs-Campbell. Traversing supervisor problem: An approximately optimal approach to multi-robot assistance. In *Robotics: Science and Systems XVIII*. Robotics: Science and Systems Foundation, 2022. ISBN 978-0-9923747-8-5. doi: 10.15607/RSS.2022.XVIII.059. URL http://www.roboticsproceedings.org/rss18/p059.pdf.

[39] Tianchen Ji, Roy Dong, and Katherine Driggs-Campbell. Traversing supervisor problem: An approximately optimal approach to multi-robot assistance. In *Robotics: Science and Systems (RSS)*, 2022.

[40] Asher Koriat. When are two heads better than one and why? *Science*, 336(6079):360–362, 2012.

[41] Asher Koriat. When two heads are better than one and when they can be worse: The amplification hypothesis. *Journal of Experimental Psychology: General*, 144(5): 934–950, 2015. doi: 10.1037/xge0000092.

[42] Ralf HJM Kurvers, Stefan M Herzog, Ralph Hertwig, Jens Krause, Patricia A Carney, Andy Bogart, Giuseppe Argenziano, Iris Zalaudek, and Max Wolf. Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, 113 (31):8777–8782, 2016.

[43] Kwang-Hyun Lee, Usman Mehmood, and Jee-Hwan Ryu. Development of the human interactive autonomy for the shared teleoperation of mobile robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[44] HCCH Levitt. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, 49(2B):467–477, 1971.

[45] Quanyi Li, Zhenghao Peng, and Bolei Zhou. Efficient learning of safe driving policy via human-ai copilot optimization. (arXiv:2202.10341), 2022. URL http://arxiv.org/abs/2202.10341.

[46] Yanan Li, Keng Peng Tee, Wei Liang Chan, Rui Yan, Yuanwei Chua, and Dilip Kumar Limbu. Role adaptation of human and robot in collaborative tasks. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015.

[47] Yuan Liu, Glenda Caldwell, Markus Rittenbruch, Müge Belek Fialho Teixeira, Alan Burden, and Matthias Guertler. What affects human decision making in human–robot collaboration?: A scoping review. *Robotics*, 13, 2024.

[48] Andrea Lockerd and Cynthia Breazeal. Tutelage and socially guided robot learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2004.

[49] Ali Mahmoodi, Dan Bang, Majid Nili Ahmadabadi, and Bahador Bahrami. Learning to make collective decisions: The impact of confidence escalation. *PLoS ONE*, 8(12): e81195, 2013. doi: 10.1371/journal.pone.0081195.

[50] Sébastien Massoni and Nicolas Roux. Optimal group decision: A matter of confidence calibration. *Journal of Mathematical Psychology*, 79:121–130, 2017.

[51] Lauren Milliken and Geoffrey A Hollinger. Modeling user expertise for choosing levels of shared autonomy. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017.

[52] A. Monferrer and D. Bonyuet. Cooperative robot teleoperation through virtual reality interfaces. In *Proceedings Sixth International Conference on Information Visualisation*, 2002. doi: 10.1109/IV.2002.1028783.

[53] Robin R Murphy. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2): 138–153, 2004.

[54] Selma Musić, Gionata Salvietti, Francesco Chinello, Domenico Prattichizzo, Sandra Hirche, et al. Robot team teleoperation for cooperative manipulation using wearable haptics. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017.

[55] Keiji Nagatani, Seiga Kiribayashi, Yoshito Okada, Kazuki Otake, Kazuya Yoshida, Satoshi Tadokoro, Takeshi Nishimura, Tomoaki Yoshida, Eiji Koyanagi, Mineo Fukushima, et al. Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots. *Journal of Field Robotics*, 30(1):44–63, 2013.

[56] Changjoo Nam, Phillip Walker, Huao Li, Michael Lewis, and Katia Sycara. Models of trust in human control of swarms with varied levels of autonomy. *IEEE Transactions on Human-Machine Systems*, 50(3):194–204, 2019.

[57] Duc-An Nguyen, Jude Nwadiuto, Hiroyuki Okuda, and Tatsuya Suzuki. Model structure identification of hybrid dynamical systems based on unsupervised clustering and variable selection. *IFAC-PapersOnLine*, 53(2):1090–1095, 2020.

[58] Duc-An Nguyen, Jude Nwadiuto, Hiroyuki Okuda, and Tatsuya Suzuki. Modeling car-following behavior in downtown area based on unsupervised clustering and variable selection method. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3714–3720. IEEE, 2020.

[59] Niccolo Pescetelli, Geraint Rees, and Bahador Bahrami. The perceptual and social components of metacognition. *Journal of Experimental Psychology: General*, 145(8): 949–965, 2016. doi: 10.1037/xge0000180.

[60] Thi Minh Anh Pham, An Duc Nguyen, Cephas Svosve, Vasileios Argyriou, and Georgios Tzimiropoulos. Pre: Vision-language prompt learning with reparameterization encoder. *arXiv preprint arXiv:2309.07760*, 2023.

[61] Thi Minh Anh Pham, Duc-An Nguyen, Cephas Svosve, Vasileios Argyriou, and Georgios Tzimiropoulos. Pre: Vision-language prompt learning with reparameterization encoder. In *International Conference on Learning Representations (ICLR 2024)*, 2024.

[62] Kyle B Reed and Michael A Peshkin. Physical collaboration of human-human and human-robot teams. *IEEE Transactions on Haptics*, 1(2):108–120, 2008.

[63] O Reinoso, Arturo Gil, Luis Payá, and Miguel Juliá. Mechanisms for collaborative teleoperation with a team of cooperative robots. *Industrial Robot: An International Journal*, 35(1):27–36, 2008.

[64] Marion Rouault, Andrew McWilliams, Micah G. Allen, and Stephen M. Fleming. Human metacognition across domains: Insights from individual differences and neuroimaging. *Personality Neuroscience*, 1:e17, 2018. doi: 10.1017/pen.2018.16.

[65] Marion Rouault, Tricia Seow, Claire M Gillan, and Stephen M Fleming. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological psychiatry*, 84(6): 443–451, 2018.

[66] Marion Rouault, Peter Dayan, and Stephen M Fleming. Forming global estimates of self-performance from local confidence. *Nature communications*, 10(1):1141, 2019.

[67] Kenji Sakita, Koichi Ogawara, Shinji Murakami, Kentaro Kawamura, and Katsushi Ikeuchi. Flexible cooperation between human and robot by interpreting human intention from gaze information. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2004.

[68] Jean Scholtz, Jeff Young, Jill L Drury, and Holly A Yanco. Evaluation of human-robot interaction awareness in search and rescue. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 3. IEEE, 2004.

[69] Jian Shen, Javier Ibanez-Guzman, Teck Chew Ng, and Boon Seng Chew. A collaborative-shared control system with safe obstacle avoidance capability. In *IEEE Conference on Robotics, Automation and Mechatronics (ICARM).*, volume 1, pages 119–123. IEEE, 2004.

[70] Sichao Song, Jun Baba, Junya Nakanishi, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Costume vs. wizard of oz vs. telepresence: how social presence forms of tele-operated robots influence customer behavior. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022.

[71] Brett Stoll, Samantha Reig, Lucy He, Ian Kaplan, Malte F. Jung, and Susan R. Fussell. "wait, can you move the robot?": Examining telepresence robot use in collaborative teams. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018.

[72] Justin Storms, Steven Vozar, and Dawn Tilbury. Predicting human performance during teleoperation. In

ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2014.

[73] Kazuaki Tanaka, Naomi Yamashita, Hideyuki Nakanishi, and Hiroshi Ishiguro. Teleoperated or autonomous?: How to produce a robot operator's pseudo presence in hri. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016. doi: 10.1109/HRI.2016.7451744.

[74] J Gregory Trafton, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 35(4):460–470, 2005.

[75] Katherine M. Tsui, Munjal Desai, Holly A. Yanco, and Chris Uhlik. Exploring use cases for telepresence robots. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011.

[76] Katherine M. Tsui, Stephen Von Rump, Hiroshi Ishiguro, Leila Takayama, and Peter Vicars. Robots in the loop: Telepresence robots in everyday life. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012.

[77] Marynel Vázquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson. Towards robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017.

[78] Michael E. Walker, Hooman Hedayati, and Daniel Szafir. Robot teleoperation with augmented reality virtual surrogates. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019. doi: 10.1109/HRI.2019.8673306.