UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
The Relationship Between Teaming Behaviours and Joint Capacity of Hybrid Human-Machine Teams

Permalink
https://escholarship.org/uc/item/61b5c7gv

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Hedley, Laiton G
Bennett, Murray S
Love, Jonathon
et al.

Publication Date
2023

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

The Relationship Between Teaming Behaviours and Joint Capacity of Hybrid Human-Machine Teams

Laiton Hedley [1], Murray S. Bennett[1,2], Jonathon Love[1], Joseph W. Houpt[2], Scott D. Brown[1], and Ami Eidels[1]

School of Psychological Science, University of Newcastle, Australia
Callaghan, NSW 2308 AUS[1]

School of Psychology, University of Texas at San Antonio [2]

Abstract

Collaboration in shared environments requires human agents to coordinate their behaviour according to the machines' actions. In this study, we compared the performance and behaviour of Human-Machine (HM) and Human-Human (HH) teams. While HH teaming behaviour is sensitive to Collaborative contexts, little is known about HM teaming behaviour. Furthermore, teaming behaviour may impact the team's *Joint Capacity* – the team's ability to handle teamwork processes and task demands. To assess teaming behaviour at every moment of a trial we used three distinct spatiotemporal measures (Momentary Distance, Highly Correlated Segments, and Running Correlation). To assess the team's joint performance, we adopted the Capacity Coefficient (Townsend & Nozawa, 1995). For both HH and HM teams, behavioural measures predicted Joint Capacity. HH teams demonstrated greater performance and less synchronous behaviour than HM teams. The reduced synchrony of HH teams likely improved their performance as they could complement each other's behaviour rather than duplicate inefficiencies.

Keywords: Human-Machine Teams; Human-Human Teams; Group performance; Collaboration; Competition; Workload Capacity; Dynamic Behavioural Measures.

Introduction

The rise in technology has led Artificial Intelligence (AI) or Machine Learning agents to work alongside human co-actors in shared environments that require coordinated behaviour amongst the human and artificial agent (or rather the *hybrid* co-actors; Innocente & Grasso, 2018) . For successful task performance, the human operators must adjust their actions according to the behaviour of their AI counterparts (Sebanz, Bekkering, & Knoblich, 2006). In our study, we focus on *implicit coordination* (when team members anticipate and integrate co-actor actions in order to adjust their own behaviour without the explicit ability to communicate) (Rico, Sánchez-Manzanares, Gil, & Gibson, 2008). The additional teamwork processes of coordination may detriment the human operator's cognitive resources, which could otherwise be used to perform the main task (Funke, Knott, Salas, Pavlas, & Strang, 2012; Young, Brookhuis, Wickens, & Hancock, 2015). Understanding this requires objective measures that can capture the underlying behavioural patterns that impact the performance of teams.

In our study, participants performed a two-player arcade-style game, "Team Spirit". Each player moved a paddle horizontally to maximise ball deflections (Figure 1). Human participants either performed alongside another human (HH team) or a machine partner (HM team). Using this task, we compared the behavioural strategies adopted by HH and HM teams. We also investigated the relationship of these behaviours with the team's ability to efficiently deal with the strain of teamwork processes and main task demands (which we call *Joint Capacity*). To assess *behavioural patterns*, we utilised state-of-the-art dynamic behavioural measures that are both spatially and temporally sensitive. We also assessed teams' *performance efficiency* by estimating the *Capacity Coefficient*. This is a measure of workload capacity typically used to asses how one individual processes multiple signals, which we scaled up to assess the team's Joint Capacity (Algom, Eidels, Hawkins, Jefferson, & Townsend, 2015; Townsend & Nozawa, 1995). We also investigate behavioural patterns known to emerge in HH teams under different social contexts. Additionally, we highlight the need to establish both the behavioural patterns adopted by HM teams and empirical evidence for the relationship between team behaviour on the team's Joint Capacity.

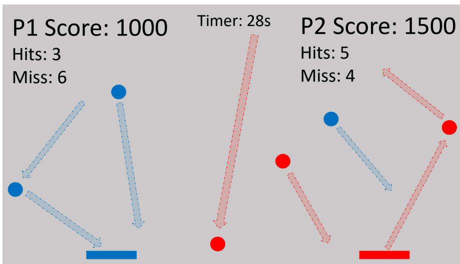

Figure 1. Illustration of the Team Spirit game. Players can move their paddles left or right to deflect balls. The arrows illustrate ball motion; they do not appear on the player's screen. In the Separate condition shown here, the red (blue) player can hit only red (blue) balls. In the group conditions (not shown), all balls are purple, and any player can hit any ball.

Teaming Environments, Behavioural Strategies, and Joint Capacity

Human teams exhibit different behavioural strategies under collaborative and competitive contexts, as found in visual search tasks in which participants used shared-gaze technology (allowing teammates to see co-actor viewing positions and behaviours). S. E. Brennan, Chen, Dickinson, Neider,

[1]Corresponding author: please address any enquiries to laiton.hedley@uon.edu.au

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*. ©2023 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

and Zelinsky (2008) found collaborative team members established a virtual boundary in the search display. This *spatial division strategy* can help team members reduce redundant search behaviours. In a similar task, Niehorster, Cornelissen, Holmqvist, and Hooge (2019) found that competitive teams showed less predictable behavioural patterns and less bias towards a particular region in the search space (as team members were likely determining their actions according to co-actor behaviour), indicating the use of a *monitoring strategy*. Increased monitoring has been linked to greater synchronization (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007) and poorer overall team performance (Yamani, Neider, Kramer, & McCarley, 2017). Our Team Spirit design shares a similar feature to these visual search tasks; the ability to view co-actor behaviour. This ability may lead to similar patterns of behaviour in other team-based tasks (i.e. Andrade-Lotero and Goldstone (2021)) and our own Team Spirit design. Our design also includes manipulations of task workload. An increase in task difficulty has led to mixed evidence on the team performance (M. C. Wright & Kaber, 2005; Proaps & Bliss, 2010) and may lead non-communicating team members to compete rather than collaborate (A. A. Brennan & Enns, 2015).

There are fewer studies focusing on hybrid HM teams in collaborative vs competitive conditions. In collaborative settings, HM teams are outperformed by HH teams (Demir, McNeese, & Cooke, 2017; Demir, Cooke, & Amazeen, 2018; Grimm, Demir, Gorman, & Cooke, 2018; McNeese, Demir, Cooke, & She, 2021). In a classification task, Steyvers, Tejeda, Kerrigan, and Smyth (2022) found that HM teams outperform HH teams when they have complementary rather than overlapping abilities. In dynamic environments, synchronous coordination patterns across team members may reduce performance since agents duplicate inefficiencies rather than complement each other. To test this, we must choose coordination measures appropriate for these dynamic task environments. However, the choice of coordination measures is dependent on the task environment, task context (such as the ability to explicitly or implicitly communicate), modalities (such as motion), and the number of team members (Halgas et al., 2023; Strang, Funke, Russell, Dukes, & Middendorf, 2014). In this paper, we explore (i) dynamic coordination measures that can capture spatiotemporal aspects of teaming behaviour in a shared environment, (ii) the behavioural differences adopted within Collaborative vs Competitive teams and (iii) the behavioural differences between dyadic teams of HH and HM co-actors.

Previous investigations have speculated consistent and predictable co-actor behaviour may reduce cognitive resources required to anticipate and integrate co-actor actions, which can improve main task performance (S. E. Brennan et al., 2008). Consistently behaving teammates have been shown to lead co-actors to adopt complementary behaviours and in turn, facilitate task performance (Andrade-Lotero & Goldstone, 2021). As for HM teams, pairing a machine teammate with a human partner has shown mixed effects on the cognitive resources of the human teammates (Azhar & Sklar, 2017; Chen & Barnes, 2012; J. L. Wright, Chen, & Barnes, 2018). Bansal et al. (2019) have established a team performance cost when there is an incompatibility between an AI's behaviour and the human co-actor's expectations of that behaviour. The existing literature highlights the possible relationship between the behavioural strategies of teams (HH and HM) and their Joint Capacity. In this study, we investigated the relationship between coordination patterns of teams and their performance.

Overview of the Experiment

In our Team Spirit arcade-game, we compared team types (HH vs HM teams), different group instructions (Collaborative, Competitive, and Separate), and workload (number of balls). We measured performance (each team's Joint Capacity) and behavioural patterns (the position of paddles at each moment) to examine (i) what behavioural strategies are adopted by HM teams and how these strategies differ from HH teams? and (ii) whether the team behaviour predicts the team's Joint Capacity, and how?

Methods

We report the methods necessary to understand the analyses of behavioural patterns. For comprehensive methods and performance analyses see Bennett et al. (submitted).

Participants

We recruited 138 participants to the HH condition and 296 to the HM team condition via Prolific and were compensated £5. After removing participants who performed poorly in the main task, there were 126 participants left in the HH condition (or 63 dyads, mean age = 34.8, SD = 11.4) and 288 participants in the HM condition (mean age = 41.8, SD = 13.0). This research was approved by the Human Research and Ethics Committee at the University of Newcastle, Australia.

Design

Team type (HH vs HM) was manipulated between subjects. We used within-subjects designs for Group Instruction (Collaborative, Competitive, and Separate) and Workload (Low, Medium, High, and Very High). Group Instruction was manipulated via the instructions to participants: Collaborative: "work together to maximise team score", Competitive: "outscore the opponent", and Separate: "score as high as possible". In the Separate condition, half the balls were red and half blue, and participants could interact only with balls that matched their paddle's colour (unmatched balls would pass through the unmatched paddle; Figure 1). In this condition, participants were each presented with the number of their individual hits (ball deflections) and misses (failure to deflect a ball). In the Collaborative and Competitive conditions, all balls were coloured purple, and participants could interact with any ball. In the collaborative condition, the players' hits were combined into a shared group score. In the competitive condition, players tried to obtain a higher individual score. We manipulated Workload by varying the number of balls per player to deflect (Low: one ball, Medium: three balls, High: six balls, and Very High: nine balls). An example of the Medium condition can be seen in Figure 1. We recorded the miss rate (for the estimation of Joint Capacity) and the position of players during each moment of the trial (subjected to behavioural analyses).

Procedure

Participants accessed the experiment online via Prolific. Right-handed participants were instructed to move the paddle left and right with the corresponding arrow keys. Left-handed participants used the z and x keys for left and right movements, respectively. Excluding practice trials, participants performed 36 trials overall. Participants performed two practice trials (without feedback) lasting 25 seconds (s) each at a low workload level and then one practice trial lasting 45s at a medium workload level. Participants completed three experimental blocks, each dedicated to the three Group Instruction conditions. Blocks were counterbalanced to avoid practice effects. Each block contained 12 trials (a trial lasted 45s). In a block, the four Workload conditions were presented three times each (resulting in a total of 12 trials per block; the order of workload trials was random for each block).

In the HH condition, two human participants were assigned to play together. In the HM condition, a human player was assigned to play with a machine agent, either a Reinforcement Learning agent or an Ideal Observer agent (see Bennett et al. (submitted)). The display in which paddles could move spanned between 40 pixels (px) and 760px. At the beginning of each trial, the paddles were positioned at 280px and 560px. A 45s timer started, and balls commenced an upward, random trajectory from the bottom of the screen (details in Bennett et al. (submitted)). Participants then tried to maximize the number of balls deflected before the trial ended. After each trial, players took a self-paced break (minimum 5s). During trial breaks, the score for the previous trial and cumulative block score was presented alongside Group Instruction reminders. After all 12 trials for a given block were completed, the participants took a minimum 20-second break and were given instructions for the next block of trials. The overall experiment lasted approximately 40 minutes.

Analyses

Analysis was conducted in JASP version 0.16.2 and MATLAB R2021b. We used several 2x2x4 mixed ANOVAs with the between-subjects variable of Team Type (HH vs HM) and within-subjects variables of Group Instruction (Collaborative and Competitive) and Workload (Low, Medium, High, and Very High). The separate condition was used only as a benchmark for our measure of Joint Capacity and was not analysed. We also conducted Linear Regression Analyses to investigate behaviour as a predictor of Joint Capacity. We checked the appropriate assumptions for all parametric analyses, Greenhouse-Giesser corrections were applied to account for violations of homogeneity of variance and Bonferroni corrections were used to account for multiple comparisons.

Joint Capacity We assessed the team's Joint Capacity using the capacity coefficient, *Cp*. This index allows us to classify the Joint Capacity of the team as either *limited*, *super* or *unlimited* (Algom et al., 2015; Townsend & Nozawa, 1995), relative to a benchmark of and Unlimited Capacity Independent Parallel (UCIP) model. The UCIP Model assumes that the performance of one co-actor is unaffected by another (Algom et al., 2015). The prediction of the UCIP model is derived as the product of the two team member's performance when performing the task alone (in the Separate condition, either player A alone or B alone). We compare this product to the team's observed performance, AB (separately for the Collaborative and Competitive conditions).

$$C(p) = p(A) \times p(B) - p(AB)$$

Where p(A) and p(B) represent the miss rates of Players A and B, and their product is the UCIP prediction. p(AB) is the combined miss rate of players A and B under conditions where they can work together (the Collaborative and Competitive conditions). *Cp*=0 implies unlimited capacity, where teamwork processes were neither detrimental nor beneficial to performance). *Cp*$<$0 implies limited capacity, where teamwork processes impacted performance negatively. *Cp*$>$0 indicates super capacity, where teamwork processes benefited the team's performance relative to the independent-processes benchmark (see Heathcote et al. (2015) for a detailed explanation of *Cp*).

Dynamic Behaviour To characterise spatiotemporal aspects of behaviour, we calculated *Momentary distance* between the two players' paddle positions, *proportion of Highly Correlated Segments* (HCS), and *Running Correlation*. The Momentary distance and HCS measures were subjected to 2x2x4 mixed ANOVA. We subjected the Running Correlation measure to several two-sample KS tests. Momentary Distance is defined as the absolute difference in the inter-paddle distance taken as the proportion of screen width during each and every moment of the game. HCS analysis measures the proportion of time segments in which the team had highly correlated or synchronous behaviour (Marcelino et al., 2020). Higher proportions indicate more synchronous behaviour. To calculate HCS proportions, we performed a cross-correlation on the paddle positions of each player, looking forwards and backwards for a given time window of 1.5s at each moment in the game. When the absolute correlation was above a threshold of 0.97, this segment was classified as a HCS. This process is repeated until all paddle positions within the moving time window at each moment or time segment of the trial have been cross-correlated and compared to the threshold. The HCS is just the proportion of segments classified as highly correlated. The Running Correlation analysis correlated the positions of player at each segment of a trial, looking forwards and backwards within a time window of 1.5s. This provided a correlation value for that moment in the trial and was repeated for each and every moment (excluding scenarios when both paddles were stationary), resulting in the frequency of correlation values for each trial and condition. We then plot histograms of the running correlation values to assess frequency. When the correlation frequencies are skewed positive or negative, the two player's paddle movements are highly synchronous or asynchronous, respectively (see Corbetta and Thelen (1996)). To investigate dynamic behaviour as a predictor of Joint Capacity we conducted Linear Regression analyses using HCS proportions and Momentary Distance as predictors of *Cp* Scores for both Team Types (HH and HM) under both the Collaborative and Competitive conditions.

Results

Momentary Distance

Figure 2 depicts the Momentary distance for both Team types. Overall, Momentary distance for both Team types is higher in the Collaborative condition than the Competitive. We found overwhelming evidence for the main effect of Group Instruction ($BF_{inclusion} > 1,000$) and a significant main effect for both Team type ($F_{(1,253)} = 39.75, p < 0.001$, $BF_{inclusion} > 1,000$) and Workload ($F_{(1.86,471.69)} = 533.66, p < 0.001$, $BF_{inclusion} > 1,000$). There were significant two-way interactions for Team type by Group Instruction, and Workload ($BF_{inclusion} > 1,000$, and ($F_{(1.9,471.69)} = 8.56, p < 0.001$, $BF_{inclusion} > 1,000$, respectively).

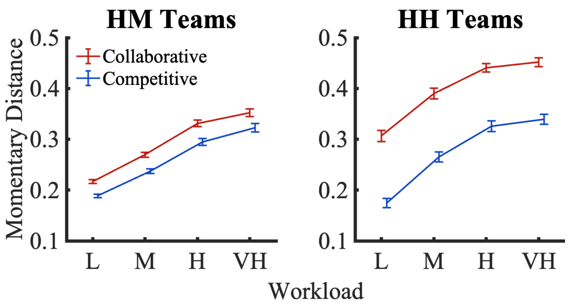

Figure 2. Momentary distance (y-axis) of HM (left) and HH teams (right) as a function of Workload (Low, Medium, High, Very High; x-axis). The individual lines represent Collaborative and Competitive conditions (with standard error bars).

HH teams had on average the greater Momentary Distance over HM teams ($t = -6.31, p < 0.001$, posterior odds $> 1,000$). Within HH teams, Momentary Distance was greater in the Collaborative condition than the Competitive (t = 17.79, $p < 0.001$). HM Team types demonstrated a similar effect (t = 8.12, $p < 0.001$). However, the increase in Momentary Distance under the Collaborative over the Competitive condition was greater for HH teams than HM teams (Mean Difference = 0.12 and Mean Difference = 0.03, respectively). On average Momentary Distance increased for both Team types as Workload levels increased (all $p < 0.001$), except for HH teams under the High condition compared to themselves under the Very High condition (t = -1.79, p = 1).

Highly Correlated Segments

Figure 3 shows the proportion of HCS for each Team type. The main effects of Team type and Workload were significant (both $p < 0.001$ and $BF_{inclusion} > 1,000$) and there was overwhelming evidence for a main effect of Group Instruction ($BF_{inclusion} > 1,000$). There were significant two-way interactions for Team Type with Group Instruction and Workload ($BF_{inclusion} > 1,000$ and $F_{(2.4,606.86)} = 28.751, p < 0.001$, $BF_{inclusion} > 1,000$, respectively).

For HM teams, there was no significant difference between mean HCS proportions in the Collaborative condition compared to the Competitive (t = 0.52, p = 1). As for the HH teams, mean HCS proportion was lower in the Collaborative condition than the Competitive condition (t = -6.29, $p < 0.001$; see Figure 3). For both HH and HM Team types, as Workload levels increased the proportion of HCS decreased (all $p < 0.001$), except for the High vs Very High conditions ($t = 0.97$, $p = 1$ and $t = 2.54$, $p = 0.315$, respectively). When comparing matched Workload levels across Team types, we found HH teams had significantly lower HCS proportions than HM teams for all Workload conditions (all $p < 0.05$) but there was no difference found between Team types in the Low conditions (t = -1.19, p = 1). Overall, HH teams have on average lower proportions of HCS than HM teams ($t = 4.17, p < 0.001$, posterior odds $> 1,000$).

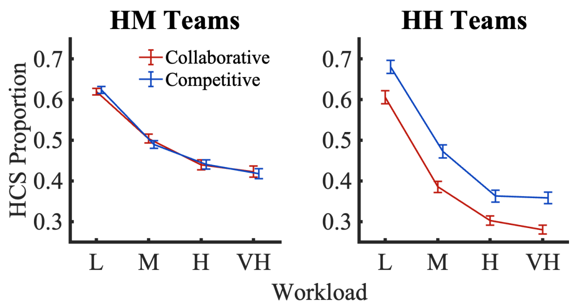

Figure 3. HCS proportions (y-axis) for HM (left) and HH teams (right), for Collaborative and Competitive conditions (individual lines with standard error bars) as a function of Workload (Low, Medium, High, Very High; x-axis).

Running Correlation

Using a two-sample K-S test we found a significant difference between the distribution of Correlation Frequencies for the two Team types ($D_{(63,191)} = 0.04$ $p < 0.001$), Correlation Frequencies appear more positively skewed for HM teams than HH teams (see Figure 4. When examining only the HH team, we found a significant difference in their distribution of Correlation Frequencies under the Collaborative condition compared to the Competitive condition ($D_{(63,63)} = 0.12$, $p < 0.001$). When comparing the effects of Collaboration vs

Competition for only HM teams, we also found a significant difference in correlation distributions between the Group Instruction conditions ($D_{(192,192)} = 0.02$, $p < 0.001$). The HM teams had a greater frequency of positive correlations in the Competitive condition than the Collaborative (see Figure 4). The distribution of correlations was more positively skewed for HM teams than HH teams in the Collaborative condition, and this difference was significant ($D_{(63,192)} = 0.09$, $p < 0.001$). Additionally, HM teams also demonstrated more heavily skewed correlation distributions than HH teams under the effect of Competition (see Figure 4 and this difference was also significant ($D_{(63,192)} = 0.05$, $p < 0.001$).

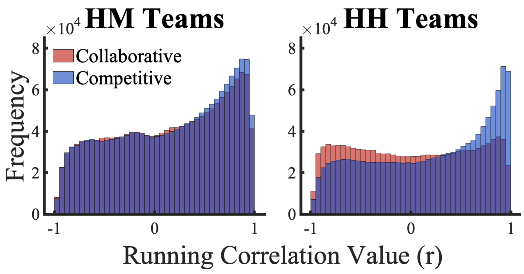

Figure 4. The Frequency (y-axis) of Running Correlation Values or r (x-axis) for HM (left) and HH teams (right).

Joint Capacity *Cp* Scores

For both Team types, *Cp* Scores were below the UCIP model, indicating the teams were performing at limited capacity (see Figure 5). We found significant main effects of Team type and Workload ($F_{(1,239)} = 126.37, p < 0.001$, $BF_{inclusion} > 1{,}000$ and $F_{(1.94,999.5)} = 168.5, p < 0.001$, $BF_{inclusion} > 1{,}000$, respectively) and overwhelming evidence for the main effect of Group Instructions ($BF_{inclusion} > 1{,}000$). There were significant two-way interactions between Team Type and Group Instruction ($BF_{inclusion} > 1{,}000$) and Team Type and Workload ($F_{(2.86,999.49)} = 4.64, p = 0.004$, $BF_{inclusion} = 154.88$).

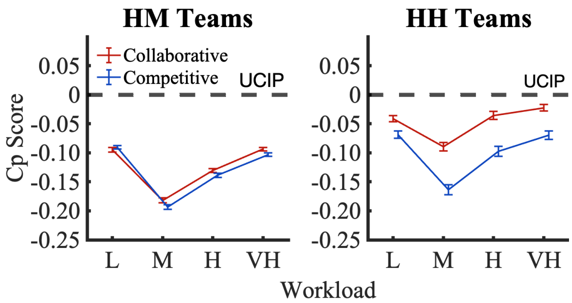

Figure 5. *Cp* Scores (y-axis) for HM (left) and HH teams (right), across the Collaborative and Competitive instructions (with standard error bars) as a function of Workload (Low, Medium, High, Very High; x-axis). The UCIP Model is represented by the dashed line.

Figure 5 shows that the main effect of Team Type on capacity is driven by the Collaborative advantage in HH teams, whereas this effect was much smaller for HM. HH teams had higher *Cp* Scores in the Collaborative condition than the Competitive (t = 11.38, $p < 0.001$; Mean Difference = 0.05). HM teams also demonstrated greater *Cp* Scores in the Collaborative condition than the Competitive (t = 2.91, p = 0.023) but this difference was small (Mean Difference = 0.01). As for differences between Team types, Collaboration and Competition resulted in greater *Cp* Scores for HH teams under both conditions (t = -14.17, $p < 0.001$ and t = -5.76, $p < 0.001$, respectively).

For both Team types *Cp* Scores did not change monotonically with load; rather, they were significantly lower under the Medium Workload condition compared to all other Workload conditions (all $ps < 0.001$). Comparing HH and HM teams, for all matched Workload comparisons HH teams had greater *Cp* Scores than HM teams (all $ps < 0.001$).

Relationships between Behaviour and Performance

Momentary Distance as a predictor of *Cp* Scores As Momentary Distance increased *Cp* Scores improved (Figure 6). Momentary Distance in the Collaborative and Competitive conditions for HH teams was a significant predictor of *Cp* Scores ($F_{(1,61)} = 28.71, p < 0.001$, $BF_{inclusion} > 1{,}000$, $R^2 = 0.31$ and $F_{(1,61)} = 40.82, p < 0.001$, $BF_{inclusion} > 1{,}000$, $R^2 = 0.39$ respectively). For HM teams in the Collaborative and Competitive conditions, Momentary Distance was also a significant predictor of *Cp* Scores ($F_{(1,191)} = 123.54, p < 0.001$, $BF_{inclusion} > 1{,}000$, $R^2 = 0.39$ and $F_{(1,191)} = 113.18, p < 0.001$, $BF_{inclusion} > 1{,}000$, $R^2 = 0.37$, respectively).

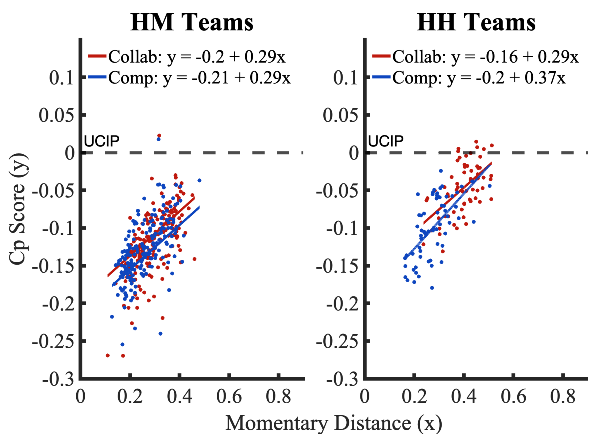

Figure 6. Momentary Distance (x-axis) as a predictor of *Cp* Scores (y-axis) for both HM (left) and HH teams (right). The UCIP model is represented by the dashed line.

HCS proportions as a predictor of *Cp* Scores As HCS proportions increased *Cp* Scores decreased (Figure 7). For HH teams HCS proportions were a significant predictor of *Cp* scores in both the Collaborative ($F_{(1,61)} = 5.78, p = 0.019$, $BF_{inclusion} = 2.79$) and Competitive conditions ($F_{(1,61)} = 14.44, p < 0.001$, $BF_{inclusion} = 77.95$), with adjusted $R^2 =$

0.07 and $R^2 = 0.18$, respectively. For HM teams HCS proportions were a significant predictor of *Cp* Scores for the Collaborative and Competitive conditions ($F_{(1,190)} = 6.95, p = 0.009$, $BF_{inclusion} = 3.87$ and $F_{(1,190)} = 6.96, p = 0.009$, $BF_{inclusion} = 3.89$, respectively). Variance in *Cp* Scores explained by HCS proportions in both the Collaborative (Adjusted $R^2 = 0.03$) and Competitive conditions (Adjusted $R^2 = 0.03$) was even smaller. The regression slope for HH teams was lower in the Collaborative condition than the Competitive and practically similar across these conditions for HM teams (Figure 7).

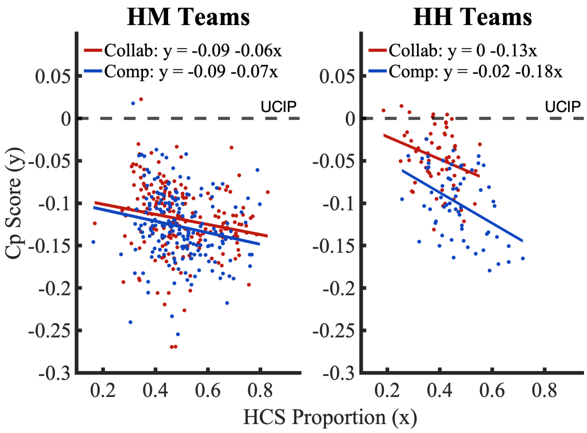

Figure 7. HCS proportions (x-axis) as a predictor of *Cp* Scores (y-axis) for both HM (left) and HH teams (right). The UCIP model is represented by the dashed line.

Discussion

We compared differences in the behaviour of HH and HM dyads. We used three spatiotemporal measures of behaviour: HCS proportions, Momentary Distance, and Running Correlation alongside the Capacity Coefficient. The within-subjects effect of Collaboration vs Competition was much greater for HH than HM teams in both teaming behaviour and Joint Capacity. Collaboration led to greater spatial separation (larger inter-paddle distance) and less synchronous behaviour compared to the Competitive setting for both Team Types. This difference was much greater for HH than HM teams. As workload levels increased, spatial division increased and synchronous behaviour decreased amongst team members for both Team types. Overall, HM teams demonstrated more synchronous behaviour and smaller inter-paddle distance than HH teams. All teams demonstrated limited capacity performance on average, meaning there were inefficiencies in their ability to handle both task demands and teamwork processes. Collaboration led to greater Joint Capacity for both Team Types over Competition, but the improvement in Joint Capacity was greater for HH teams than HM teams.

Our results suggest a relationship between teaming behaviour and Capacity. Synchronous behaviour increases the cost to Joint Capacity. HH teams displayed a reduction in this cost in the Collaborative condition (compared to the Competitive) while HM teams did not. For both Team types, as inter-paddle distance increases Joint Capacity improves. Our study also demonstrated the sensitivity of HH team's behaviour to the social context of the teaming environment and the lack of such sensitivity in hybrid HM teams.

Our results also demonstrate an important difference in the Joint Capacity of HH and HM teams and established an empirical relationship between teaming behaviour and Joint Capacity. Human partners can reciprocate and adjust according to the co-actors (Sebanz et al., 2006). It is not surprising teams with two humans demonstrate superior abilities in handling both task demands and team coordination than hybrid HM teams (with only one team member capable of such abilities). These findings on behavioural strategies of collaborative and competitive HH teams are similar to those observed in other tasks (S. E. Brennan et al., 2008; Niehorster et al., 2019; Andrade-Lotero & Goldstone, 2021). As for the HM teams, Collaboration was not enough to elicit efficient behaviour and performance. Our artificial agents were designed to perform at a human level to ensure the benefits observed in Joint Capacity were the result of implicit coordination. Adopting machine partners with the ability to integrate teammate behaviour may improve the team's Joint Capacity, how this improvement would fare in comparison to HH teams is still an empirical question. The development of these machine agents was beyond the scope of the current study. Overall, our results demonstrate teams can experience larger benefits by taking into consideration the behaviour of co-actors (Bansal et al., 2019; Demir et al., 2017; Funke et al., 2012).

An unexpected finding was the non-monotonic relationship between load and capacity. Future research should consider non-linear relationships between teaming behaviour and performance. The behaviour analyses used in our study may also prove useful in analysing the relationships between team member eye patterns and hand gestures (Caruana, Inkley, Nalepka, Kaplan, & Richardson, 2021), or in understanding how optimal teaming behavioural patterns develop over time (Andrade-Lotero & Goldstone, 2021). The HCS analysis cannot determine if the team's behaviour is asynchronous or synchronous, as it takes the absolute correlation value, caution should be taken when interpreting HCS outcomes. For future linear behavioural analyses, such as HCS and Running Correlation, it is important to control for the possibility of spurious correlations. We have piloted Machine-Machine teams, which be used in the future as a benchmark to compare against HH and HM correlations (Strang et al., 2014).

In conclusion, our study implemented spatiotemporal measures to assess teams' behaviour. These measures predict the teams' Joint Capacity and suggest that HH behaviour is less correlated than that of HM teams. This reduced HH synchrony is associated with better performance, compared with HM teams, perhaps because it allows human players to complement each other rather than duplicate inefficiencies.

Acknowledgments

This research was supported by AUSMURI and ARC-DP grants to SB and AE (AUSMURIIV000001 and DP210100313), and by the Australian Government RTP Scholarships to LH and MB.

References

Algom, D., Eidels, A., Hawkins, R. X., Jefferson, B., & Townsend, J. T. (2015). Features of response times: Identification of cognitive mechanisms through mathematical modeling. *The Oxford handbook of computational and mathematical psychology*. doi: 10.1093/oxfordhb/9780199957996.013.4

Andrade-Lotero, E., & Goldstone, R. L. (2021). Self-organized division of cognitive labor. *PloS one*, *16*(7), e0254532. doi: 10.1371/journal.pone.0254532

Azhar, M. Q., & Sklar, E. I. (2017). A study measuring the impact of shared decision making in a human-robot team. *The International Journal of Robotics Research*, *36*(5-7), 461–482. doi: 10.1177/0018720810366859

Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., & Horvitz, E. (2019). Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 33, pp. 2429–2437). doi: 10.1609/aaai.v33i01.33012429

Bennett, M. S., Hedley, L., Love, J., Houpt, J. W., Brown, S. D., & Eidels, A. (submitted). Human performance in competitive and collaborative human-machine teams. *Topics in Cognitive Science*.

Brennan, A. A., & Enns, J. T. (2015). When two heads are better than one: Interactive versus independent benefits of collaborative cognition. *Psychonomic Bulletin & Review*, *22*(4), 1076–1082.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*(3), 1465–1477.

Caruana, N., Inkley, C., Nalepka, P., Kaplan, D. M., & Richardson, M. J. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific reports*, *11*(1), 21037.

Chen, J. Y., & Barnes, M. J. (2012). Supervisory control of multiple robots in dynamic tasking environments. *Ergonomics*, *55*(9), 1043–1058.

Corbetta, D., & Thelen, E. (1996). The developmental origins of bimanual coordination: a dynamic perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 502. doi: 10.1037/0096-1523.22.2.502

Demir, M., Cooke, N. J., & Amazeen, P. G. (2018). A conceptual model of team dynamical behaviors and performance in human-autonomy teaming. *Cognitive Systems Research*, *52*, 497–507. doi: 10.1016/j.cogsys.2018.07.029

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12. doi: 10.1016/j.cogsys.2016.11.003Get

Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and measurement of team workload: A critical need. *Human Factors*, *54*(1), 36–51.

Grimm, D. A., Demir, M., Gorman, J. C., & Cooke, N. J. (2018). Team situation awareness in human-autonomy teaming: A systems level approach. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 62, pp. 149–149). doi: 10.1177/1541931218621034

Halgas, E. A., van Eijndhoven, K. H., Gevers, J. M., Wiltshire, T. J., M. Westerink, J. H., & Rispens, S. (2023). A review of using wearable technology to assess team functioning and performance. *Small Group Research*, *54*(1), 41–76. doi: 10.1177/10464964221125717

Heathcote, A., Coleman, J. R., Eidels, A., Watson, J. M., Houpt, J., & Strayer, D. L. (2015). Working memory's workload capacity. *Memory & cognition*, *43*(7), 973–989. doi: 10.3758/s13421-015-0526-2

Innocente, M. S., & Grasso, P. (2018). Swarm of autonomous drones self-organised to fight the spread of wildfires. In *Rsff.*

Marcelino, R., Sampaio, J., Amichay, G., Gonçalves, B., Couzin, I. D., & Nagy, M. (2020). Collective movement analysis reveals coordination tactics of team players in football matches. *Chaos, Solitons & Fractals*, *138*, 109831.

McNeese, N. J., Demir, M., Cooke, N. J., & She, M. (2021). Team situation awareness and conflict: A study of human–machine teaming. *Journal of Cognitive Engineering and Decision Making*, *15*(2-3), 83–96.

Niehorster, D. C., Cornelissen, T., Holmqvist, K., & Hooge, I. (2019). Searching with and against each other: Spatiotemporal coordination of visual search behavior in collaborative and competitive settings. *Attention, Perception, & Psychophysics*, *81*(3), 666–683.

Proaps, A., & Bliss, J. (2010). Team performance as a function of task difficulty in a computer game. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 54, pp. 2413–2416). doi: 10.1177/154193121005402811

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, *26*(6), 867–891. doi: 10.1016/j.humov.2007.07.002

Rico, R., Sánchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *Academy of management review*, *33*(1), 163–184. doi: 10.2307/20159381

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70–76.

Steyvers, M., Tejeda, H., Kerrigan, G., & Smyth, P. (2022). Bayesian modeling of human–ai complementarity. *Proceedings of the National Academy of Sciences*, *119*(11), e2111547119. doi: 10.1073/pnas.2111547119

Strang, A. J., Funke, G. J., Russell, S. M., Dukes, A. W., & Middendorf, M. S. (2014). Physio-behavioral coupling in a cooperative team task: contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 145. doi: 10.1037/a0033125

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*(4), 321–359. doi: 10.1006/jmps.1995.1033

Wright, J. L., Chen, J. Y., & Barnes, M. J. (2018). Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. *Ergonomics*, *61*(8), 1033–1045. doi: 10.1080/00140139.2018.1441449

Wright, M. C., & Kaber, D. B. (2005). Effects of automation of information-processing functions on teamwork. *Human Factors*, *47*(1), 50–66. doi: 10.1518/0018720053653776

Yamani, Y., Neider, M. B., Kramer, A. F., & McCarley, J. S. (2017). Characterizing the efficiency of collaborative visual search with systems factorial technology. *Archives of Scientific Psychology*, *5*(1), 1–9. doi: 10.1037/arc0000030

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17. doi: 10.1080/00140139.2014.956151