When Collaboration Beats Ability: Mixed-Ability Teams Can Outperform High-Ability Teams Under Coordination Demands

Younes Strittmatter[1*], **Rachael Skye**[2*], **Samuel Lozano Iglesias**[3], **Samuel Liebana**[4], **Andrew Saxe**[4], **Miguel Ruiz-Garcia**[3], **Erin Teich**[2*], **Markus Spitzer**[5*]
[1]Department of Psychology, Princeton University, Princeton, NJ 08544, USA
[2] Department of Physics and Astronomy, Wellesley College, Wellesley, MA, 02481, USA
[3]Departamento de Estructura de la Materia, Física Térmica y Electrónica, and Grupo Interdisciplinar de Sistemas Complejos, Universidad Complutense Madrid, 28040 Madrid, Spain
[4]Gatsby Unit & Sainsbury Wellcome Centre, University College London, London W1T 4JG, United Kingdom
[5]Department of Psychology, Martin-Luther University Halle-Wittenberg, Halle, Germany
[*]These authors contributed equally to this work.
Corresponding author: markus.spitzer@psych.uni-halle.de

Abstract

Collective intelligence describes the capacity of groups to achieve levels of performance that cannot be explained by the abilities of their individuals. The existence of collective intelligence implies that groups composed of mixed-ability members may outperform groups of high ability. Here, we examine when such advantages emerge in humans and whether emergence is driven by collaboration. We designed a collaborative multiplayer online game and manipulated team composition (mixed-ability vs. high-ability) and coordination demands by exposing teams to two different task environments that varied in how much they encouraged collaboration. We collected data from 280 teams of two human players (70 per condition), totaling 560 participants. We found that mixed-ability teams outperformed high-ability teams in environments designed to encourage collaboration. This performance advantage was due to more collaborative actions and more frequent division of labor. Our results show that collaboration emerges selectively as a function of group composition and coordination demands.

Keywords: collective intelligence; collaboration; multiplayer online game; overcooked; ability diversity; mixed-ability

Introduction

Collective intelligence refers to the ability of groups to achieve levels of performance that cannot be fully explained by the abilities of individual members alone (Woolley et al., 2010). Collective intelligence has been documented in a wide range of settings (Kim et al., 2017; Malone & Bernstein, 2015; Riedl et al., 2021; Woolley et al., 2015), and a growing body of empirical work suggests that collective intelligence is closely related to how group members collaborate (Engel et al., 2014; Kim et al., 2017; Riedl et al., 2021; Woolley et al., 2010). Across a variety of tasks, group performance is better predicted by collaboration-related factors, such as social sensitivity, balanced participation, and communication dynamics, than by the cognitive abilities of the most capable individuals (Engel et al., 2014; Woolley et al., 2010). More recent work has further shown that the collaboration process of groups while performing a task is correlated with their task performance (Riedl et al., 2021). While these findings show that collaboration plays a central role in collective intelligence, they typically measure the degree of collaboration and its relationship to collective intelligence. As a result, they leave open a critical question: *when does collaboration actually emerge?* In other words, under what circumstances are groups more likely to engage in collaborative strategies rather than relying on individual strategies?

To address this question, experimental paradigms are needed in which collaboration is optional, dynamic, and directly observable (Janssens et al., 2022). Game-based paradigms provide a powerful approach for studying the emergence of collaboration because they embed decision-making and interaction in rich environments while maintaining experimental control (Allen et al., 2024; Brändle et al., 2023; Kim et al., 2017). In particular, multiplayer games offer a natural setting for investigating how collaborative strategies emerge from real-time interaction. For example, collaborative multiplayer cooking games inspired by *Overcooked* (Games, 2016) have previously been used to study coordination among artificial intelligence (AI) agents and among humans and AI agents (Biswas et al., 2025; Knott et al., 2021; Li et al., 2023; Strouse et al., 2022; Wu et al., 2021; Yu et al., 2023; Zhao et al., 2022). However, a systematic understanding of when and why collaboration emerges in human teams in this game environment remains largely missing.

Here, we work toward this understanding by developing an *Overcooked*-like online multiplayer game in order to examine the emergence of collaboration in human teams. We use our multiplayer online game to particularly study the emergence of collaboration under controlled variations in team composition and coordination demands. Prior work has shown that these factors, *i.e.*, the diversity of team members (Aggarwal et al., 2019; Hong & Page, 2004; Rocca & Tylén, 2022) and the extent to which coordination is required for task completion (Biswas et al., 2025; Verhagen et al., 2022), shape the collaboration and performance of teams in a variety of settings. Through systematic variation of these factors using our multiplayer online game, we collect a large dataset of team behaviors (560 participants) and use it to demonstrate that collaboration is especially likely in human teams composed of members with differing ability strengths in constrained task environments. This enhanced collaboration likelihood enables these teams to out-perform (on average) teams composed of high-ability members with a higher average ability, thereby signaling the team performance enhancement that is the hallmark of collective intelligence.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

A multiplayer online game to study collaboration

In our game, two human players control a single chef each, and all chefs within a kitchen can assemble and deliver salads to achieve a collective score (the total number of delivered salads across players). Players act by clicking on locations in the map. Clicking on items such as ingredients or plates allows players to pick them up, while clicking on kitchen counters or cutting boards places items down. Collisions are possible; *i.e.*, players cannot walk through one another or occupy the same position on the map. A successful delivery requires teams to pick up ingredients (our experiments only comprised one ingredient, tomatoes), cut them on a cutting board, place a plate on the kitchen counter, assemble the salad on the plate, and deliver it to the delivery station. Successful team performance, therefore, depends on completing two time-consuming subtasks: (a) walking through the kitchen to transport tomatoes and empty plates and to deliver assembled salads, and (b) cutting tomatoes at a cutting board.

We designed the game to allow the gradual adjustment of players' walking and cutting abilities. In addition, the game includes a set of kitchen maps that vary in how strongly they encourage coordination (Figure 1). The game thereby allows systematic investigation of collaboration under different task environments with varying coordination demands (for a similar game, see Carroll et al., 2019).

Critically, collaboration in this game can only emerge through non-verbal coordination embedded in players' actions. Players cannot talk to each other during the game and receive no instructions about which actions to take or how to coordinate, leaving them free to decide how to act throughout gameplay. Players can divide labor, with one player focusing on sub-task (a) (taking all walking-intensive actions) and one player focusing on sub-task (b) (cutting tomatoes). Alternatively, players may complete the task using an individual strategy, in which each player performs all subtasks independently. Thus, the game design allows players to apply individual and collaborative strategies within the same task.

The present study

In this study, we used our multiplayer online game to systematically vary team composition and coordination demands, and monitor team collaboration (quantified by actions where one player picked up a plate, tomato, or assembled salad that the other player set down on the kitchen counter/island) and performance (quantified by the team score).

We manipulated team composition by varying chefs' abilities. In high-ability teams, both players' chefs had maximal walking speed and maximal cutting speed. In mixed-ability teams, each chef had maximal speed in one ability but was reduced to 40% speed in the other, such that one player was a fast walker but slow cutter, while the other showed the opposite profile. Importantly, this manipulation introduced only deficits and no performance advantages: mixed-ability teams were strictly dominated by high-ability teams in terms of individual capabilities. This allowed us to test whether ability

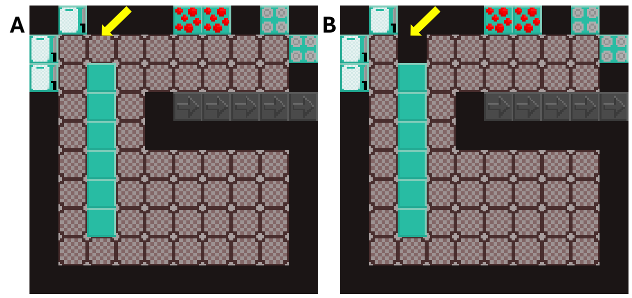

Figure 1: Game maps. A: Open map. B: Partially-blocked map. Players can walk on checkered brown tiles. Players can take plates (gray circles) from the two counter tiles corresponding to the plate dispenser in the upper right corner of the game and tomatoes (red circles) from two counter tiles next to the plate dispenser. Players can also put plates and tomatoes on the turquoise kitchen counter and cut tomatoes on the three cutting boards in the upper left corner of the game. Finally, players can deliver assembled salads on the grey tiles with arrows. Players cannot walk or place items on black tiles. The yellow arrow is not part of the game map and indicates the key difference between the two maps: players can walk through this tile on the open map, while players cannot move through this tile on the partially-blocked map.

diversity, while having the same set of abilities, promotes collaboration.

We manipulated the task environment by applying two game maps that differ in their structural layout. The *open map* (Figure 1A) does not encourage collaboration much for high-ability players, as individual players can move efficiently through the top of the map above the kitchen island and therefore perform both subtasks with minimal walking. In contrast, we designed a *partially-blocked map* (Figure 1B) to strongly encourage collaboration. This map increases walking demands substantially for players pursuing independent strategies because it requires them to walk around the kitchen counter to perform both subtasks. Dividing labor in the partially-blocked map should therefore result in substantially enhanced performance. Manipulating task environment allowed us to test how increased coordination demands, induced by environmental structure, shift the relative benefits of individual versus collaborative strategies and thereby shape whether collaboration emerges.

Method

Participants

We collected data through Prolific from 70 teams for each of four game conditions, for a total of 280 teams and 560 participants (mean age = 36.69; SD = 8.23). All participants gave consent to participate in this study and were at least 18 years old. Participation was voluntary and participants could leave the experiment at any time. The experiment lasted for 5 minutes, and we compensated participants with $1 (USD) for participating, plus a performance bonus of \$0.02 per point scored in the testing phase as a team.

Experimental design

We implemented the above-described game with two players in the 2×2 experimental design. The goal of the game was to deliver tomato salads. We did not instruct players to apply certain strategies and players could not talk to each other before or during the game, or participate in multiple games. We did not instruct participants about the abilities of their or their teammate's chef. In the mixed-ability teams, players could infer from their own observation that one player walked relatively quickly and one player cut relatively quickly. Players always had identical initial positions on the kitchen maps: one player next to the cutting boards and one player next to the plates and ingredients. On mixed-ability teams, the player with fast cutting ability started next to the cutting board while the player with fast walking ability started next to the plates. Players always saw the collective team score during the game, and we did not present individual scores to players.

The experiment consisted of two phases: a training phase (2 minutes) and a testing phase (3 minutes). Before training, we informed players that team scores did not matter during training. Before testing, we informed them that team scores affected their bonus during testing.

Data analysis

To examine how team composition and task environment influenced team performance and collaboration, we ran a 2×2 analysis of variance (ANOVA) in R. Post hoc pairwise comparisons of estimated marginal means were conducted using the *emmeans* package (Lenth, 2024), with Tukey adjustment to control for multiple comparisons. In addition, we performed behavioral clustering analyses in *Python* using the *scikit-learn* package (Pedregosa et al., 2011) and quantified differences in groups with respect to team scores and collaboration with Bonferroni corrected two-sided t-tests. Due to page limits, we report only data from the testing phase.

As independent variables, we considered two dichotomous factors: *team composition* (high-ability vs. mixed-ability teams) and *game map* (open vs. partially-blocked map). As dependent variables, we analyzed two complementary measures: (i) *team score*, defined as the total number of delivered plates per team, and (ii) *collaboration index*, defined as the proportion of collaborative actions within a team. We computed collaboration index as the number of collaborative actions (e.g., assembling a dish using tomatoes placed by the partner) divided by the total number of team actions.

For each dependent variable, we ran an ANOVA including the main effects of team composition and game map, as well as their interaction. We quantified differences between conditions with pairwise comparisons of estimated marginal means. To further examine the relationship between collaboration and performance, we correlated collaboration index and team score for each condition to assess how variation in collaboration within conditions relates to variation in team performance.

Table 1: Codes for the 10 most frequent actions

Action	Code
Use cutting board	0
Pick up ingredients from cutting board	1
Assemble salad	2
Pick up ingredients from counter	3
Put plate on counter	4
Pick up ingredients from dispenser	5
Pick up plate from dispenser	6
Pick up salad from counter	7
Put ingredients on counter	8
Deliver salad	9

To investigate how players differ in behavior, we encoded their actions during a game as a vector containing the total number of times they perform each action. Players took a total of 26 distinct types of action during a game; however, the 10 most frequent actions corresponded to 93% of all actions taken. For ease of reference, the 10 most frequent actions are numbered in Table 1. The encoded actions form a vector that represents an aggregate template of each player's choices, such that two vectors with a high difference correspond to two players with very divergent behavior.

To classify behaviors across this high-dimensional action space, action vectors were clustered using the density-based DBSCAN algorithm, which identifies similar clusters as high-density regions in the space. These groupings can be labeled as behavior archetypes and used to investigate how archetypes vary across each condition. We chose hyperparameters $\varepsilon = 4.2$ and minimum cluster size $N = 8$, leaving a total noise ratio of 38%. These parameters were chosen because they yield a robust and interpretable clustering with minimal unclassified points.

Results

We illustrate the results of our ANOVA analyses on game score and collaboration index in Fig. 2.

Game score

The ANOVA results revealed a significant main effect of team composition ($F = 6.23$, $p = .009$), indicating a higher score for high-ability teams than mixed-ability teams. There was no main effect of map ($F = 0.35$, $p = .557$). Critically, we observed a strong team composition × map interaction ($F = 66.78$, $p < .001$), indicating that the effect of group composition on performance depended on the task environment.

To follow up the significant team composition × map interaction, we conducted Tukey-adjusted pairwise comparisons of estimated marginal means. On the *open map*, high-ability teams scored significantly higher than mixed-ability teams, ($t = 7.63$, $p < .001$). In addition, high-ability teams scored significantly more points on the open map compared

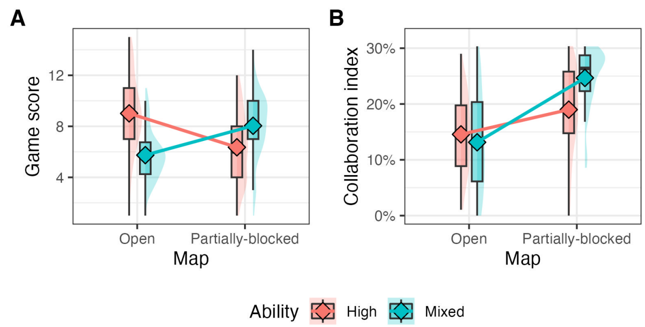

Figure 2: Differences between conditions are visualized with boxplots, averages (diamonds), and half-violin plots for (A) game score and (B) collaboration index.

to high-ability teams on the partially-blocked map ($t = 6.20$, $p < .001$). Critically, mixed-ability teams scored significantly higher than high-ability teams on the partially-blocked map ($t = 3.93$, $p < .001$). Mixed-ability teams on the partially-blocked map also scored significantly higher than mixed-ability teams on the open map ($t = -5.36$, $p < .001$).

Together, these results suggest that under the specific circumstances of our game, high-ability teams performed best in environments like the open map where individual strategies are effective, whereas mixed-ability teams can outperform high-ability teams when the task environment encourages coordination.

Collaboration index

The ANOVA results revealed significant main effects of team composition ($F = 6.32$, $p = .013$) and map ($F = 88.55$, $p < .001$), as well as a significant team composition $\times$ map interaction ($F = 17.55$, $p < .001$).

On the *open map*, collaboration did not differ between high-ability and mixed-ability teams ($t = 1.18$, $p = .637$). In contrast, on the partially-blocked map, mixed-ability teams showed significantly higher collaboration than high-ability teams ($t = 4.74$, $p < .001$). In addition, high-ability teams collaborated more on the partially-blocked map than on the open map ($t = -3.69$, $p = .002$) and the same pattern held for mixed-ability teams ($t = -9.62$, $p < .001$).

Together, these findings show that collaboration increased selectively under environmental conditions that promoted coordination, with mixed-ability teams showing the most collaboration in environments that encourage division of labor.

Correlation between collaboration and game score

Collaboration index was significantly positively correlated with team performance across all conditions (all $p < .001$), with the strongest correlation observed for mixed-ability teams and the partially-blocked map.

Classifying collaboration through action profiles

To investigate how environmental constraints mechanistically give rise to collaboration between players, we examined their behavior in more detail, paying particular attention to the set of actions each player took during gameplay. We found that players could be sorted into four groups via DBSCAN clustering of their actions (Fig. 3A), and that groups had significantly varying levels of collaboration (Fig. 3B).

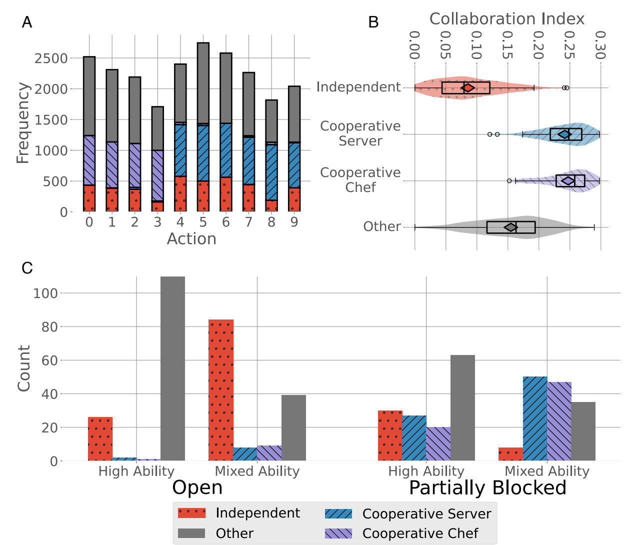

Figure 3: Player behaviors can be classified into distinct groups via DBSCAN clustering. Groups differ in (A) their relative frequency of specific actions and (B) collaboration index, visualized via boxplots, averages (diamonds), and violin plots. (C) Groups are unevenly distributed across game conditions, shown via histograms of player behavior for each condition. Note that, within each condition, all counts of player archetypes add to 140.

The first group, labeled Independent, took most common actions at equal rates except for actions 3 and 8 which players do not take when not collaborating (Fig. 3A, see Table 1 for a description of each action). The interpretation that the Independent group did not collaborate frequently is supported by a low average collaboration index in their games, ≈ 0.08 (Fig. 3B). By contrast, the two groups labeled Cooperative Chef and Cooperative Server took distinct and essentially non-overlapping sets of actions (Cooperative Chefs took actions 0 to 3 and Cooperative Server took actions 4 to 9; Fig. 3A) and participated in games with a high average collaborative index, ≈ 0.25 for both groups (Fig. 3B). The differences in collaboration index between the Independent group and the Cooperative Server group ($t = 9.54$, $p < .001$) and Independent group and the Cooperative Chef group ($t = 8.89$, $p < .001$) were statistically significant, while the Cooperative Server group and the Cooperative Chef group did not significantly differ in collaboration index ($t = 0.65$, $p = .451$).

The Cooperative Chef group tended to take actions localized on the left side of the map, around the cutting board, while the Cooperative Server group tended to take actions localized on the right side of the map, interacting with the dispensers and delivery counter (subtasks b and a of the game as described above; Fig. 3A). These behavior archetypes occur in teams; 80% of games involving a Chef or Server were a Chef/Server pair, while in 18% of games their partner was in the Other group and in only 2% of games a Chef or Server was partnered with an Independent player. Collectively, these action archetypes indicate that the Cooperative Server and Cooperative Chef players divided tasks during games, with the Cooperative Chef using the cutting board and assembling the dish and the Cooperative Server bringing ingredients to the counter and delivering dishes.

The group labeled Other encapsulates three other clusters found by the DBSCAN algorithm: two small clusters consisting of players with other collaborative behavioral profiles similar to the Cooperative Servers/Chefs, and a large unclassified group consisting of behavioral profiles that were not clustered by DBSCAN. Few players ($\approx 5\%$) were classified into the two clusters similar to the Cooperative Server and Cooperative Chef groups. Overall, players in the Other group took a more evenly distributed set of actions and varied largely in collaboration index. This group may capture players with different strategies, as well as players who swapped strategies during a game.

Importantly, the behavioral groups identified by DBSCAN are unevenly distributed across game conditions, demonstrating that environmental constraints influence player behavior (Fig. 3C). When playing on the open map, members of both mixed-ability and high-ability teams adopted the Cooperative Chef and Cooperative Server archetypes infrequently, tending instead to adopt Independent and Other behaviors. Players on high-ability teams adopted Other behaviors more frequently, and Independent archetypes less frequently, than players on mixed-ability teams. Since the Other behavior category is characterized by a broad distribution of actions, we interpret this result to mean that high-ability teams displayed more behavioral variation than mixed-ability teams.

On the partially-blocked map, the Cooperative Chef and Cooperative Server archetypes were adopted more frequently than on the open map by members of both mixed-ability and high-ability teams (Fig. 3C). On mixed-ability teams, collaboration was the dominant strategy, as more players were classified as Cooperative compared to Independent or Other. On high-ability teams, however, players were more often classified as Other than either Cooperative or Independent (Fig. 3C).

Distributions of team scores for each behavior group in each game condition (Fig. 4) show a clear link between the performance advantage of a cooperative strategy over an independent one and players' tendencies to adopt those strategies. Two-sided t-tests indicated that Independent groups (across all conditions) scored significantly less than the Cooperative Server group ($t = 11.48$, $p < .001$) and Cooperative Chef group ($t = 13.41$, $p < .001$). The Cooperative Server group and the Cooperative Chef group did not significantly differ in team score across conditions ($t = 0.791$, $p = .551$). This suggests that all teams gained an advantage by adopting a cooperative strategy over an independent one.

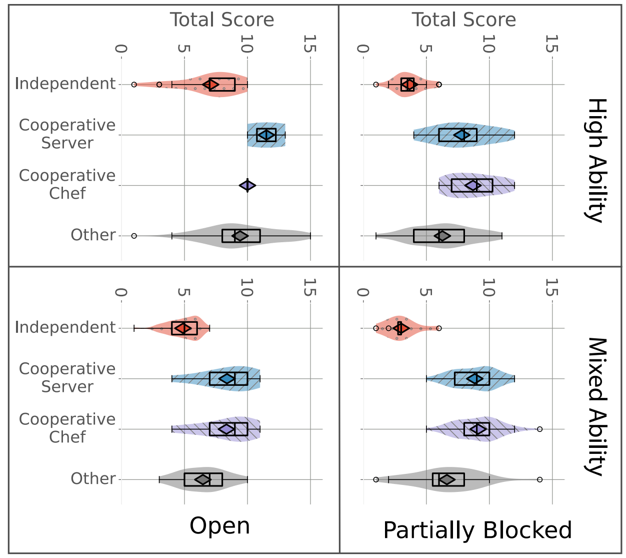

Figure 4: Team score for each behavior group in each game condition, visualized via boxplots, averages (diamonds), and violin plots. Note that for high-ability teams on the open map, very few players adopted Cooperative behaviors.

Independent players from high-ability teams had a significantly higher mean team score (≈ 7.0) on the open map compared to the partially blocked map (≈ 3.6; $t = 7.12$, $p < .001$). Similarly, independent players from mixed-ability teams had a significantly higher mean team score (≈ 4.9) on the open map compared to the partially blocked map (≈ 3.0; $t = 3.63$, $p = .006$). Especially for high-ability teams on the open map, a relatively high mean score for Independent and Other strategies (Fig. 4) leads to almost no players adopting Cooperative behaviors (Fig. 3C).

Conversely, mixed-ability teams on the partially-blocked map have a strong tendency to behave cooperatively (Fig. 3C) and a correspondingly low score when they act independently (Fig. 4). In fact, the difference in average scores between the Independent group and the two Cooperative groups (Server and Chef) was largest on the partially-blocked map for mixed-ability teams and smallest on the open map for high-ability teams[1]. This result suggests that the performance gap between the Independent and Cooperative groups may be a key factor controlling the emergence of collaboration. Notably, however, in all conditions some players in the Other category are able to achieve high scores, indicating that not all successful strategies are captured by collabo-

[1]This difference in differences was statistically significant with $p < .001$.

ration through the Chef/Server model of subdividing tasks.

In summary, our behavioral analysis aligns with our analysis of acts of collaboration, drawing a clear connection between behavior archetypes, collaboration, and performance. Players are more likely to collaborate when doing so gives a larger performance advantage, and they collaborate through task division by tacitly assigning roles to each team member despite having no verbal communication channels.

Discussion

With this study, we have systematically investigated the collaboration of human teams playing an online multiplayer game, and found that frequency of collaboration can be tuned solely through the imposition of constraints on the players. Despite having never worked together before, and despite the lack of verbal and non-verbal communication during gameplay, pairs of human players show frequent collaboration when division of labor is encouraged through two constraints: (i) asymmetry in group composition, such that players have differing ability strengths, and (ii) partial blockage in the game map such that different subsets of resources are less accessible to each player. The enhanced tendency to collaborate under these constraints enables mixed-ability teams, with lower average ability, to out-perform uniformly high-ability teams on average. Our results are consistent whether we measure collaboration by the number of acts of collaboration involving the player-to-player exchange of a shared item necessary to complete the task, or whether we determine collaboration according to behavioral archetypes of players (Cooperative Chef or Cooperative Server) based on their actions.

Collaboration in scenarios in which it is not required is far from guaranteed to emerge, even when it would improve group performance. In many tasks, individuals can achieve adequate outcomes by acting independently, creating situations of non-required cooperation (Zhang et al., 2016) in which the incentives to coordinate are ambiguous. In such settings, players who have not interacted before often converge on non-cooperative strategies, even when coordinated strategies would be more efficient (Biswas et al., 2025). Convergence to non-cooperative strategies is likely because cooperative solutions are rarely explored and remain latent—a phenomenon described in game theory as a *shadowed equilibrium* (Matignon et al., 2012). Our study sheds light on the conditions under which groups transition away from this non-cooperative equilibrium and adopt collaborative strategies. Taken together, our results provide a useful set of guidelines for enhancing the likelihood of cooperation in contexts in which it is not required.

We hypothesize that environmental constraints increase the likelihood of collaboration when they make the performance benefits of coordination sufficiently large to outweigh the costs, manifested by lower scores or time lost when a bid for collaboration by one player is ignored by the other player. Although collaboration was always an available choice for the players, they did not adopt it uniformly across conditions. Instead, teams more frequently collaborated under conditions that decreased the effectiveness of individual strategies in securing a high game score. The group most likely to collaborate was the mixed-ability team on the partially-blocked map—for this group, limitations on ability and resource access encouraged coordination in order to compensate through role differentiation.

Although our results point to the existence of a cost-benefit analysis underlying players' choices of strategy (also see Straub et al., 2023; Vélez et al., 2023), future work is needed to uncover exactly when this analysis is performed and how players adapt their behavior in real time in response to it. Our multiplayer online game will provide a means for investigating this learning in real time, since it is a fully customizable environment in which game duration can be altered and all behavior is recorded. Future studies will implement games of longer duration, possibly including reinforcement learning or other AI agents to augment the dataset, in order to elucidate the dynamics of strategy-learning in collaborative teams. Further, since the difference between mixed-ability and high-ability teams can be varied continuously in our design, future work should investigate what difference is necessary to give rise to the coordination benefits seen here.

Our work has several limitations. By its nature, our game is limited in terms of its ability to access true causal mechanisms for collaboration in human teams. For example, we cannot know the actual abilities of the human players, including their beliefs about their teammates and the resulting effects on collaboration (Xiang et al., 2023), because the game manipulates only the abilities of the chefs—not those of the human players. Additionally, while players were rewarded for a high game score, we cannot guarantee that they chose to actually maximize game score. Extensive literature shows that humans may choose to follow a personally rewarding strategy even if external incentives are clear (Yu et al., 2023). Also, we only considered one testing phase length. Our results may change with longer testing phases; for example, high-ability teams may explore new strategies and collaborate more at later times. Finally, we have focused on understanding group behavior within one task only. Others have shown that enhanced group performance due to collective intelligence is relatively stable across a range of tasks (Malone & Bernstein, 2015; Woolley et al., 2010, 2015). Future work will examine other factors that influence group behavior, and whether group behavior is consistent across a set of tasks.

To conclude, our study shows that collaboration in humans emerges selectively as a function of group composition and coordination demands. Using an online multiplayer game, we demonstrate that mixed-ability groups can outperform uniformly high-ability groups through coordination and division of labor, which they are more likely to adopt when environmental constraints increase the performance advantage of collaborative strategies. Our results point toward the strategic conditions under which collaboration becomes worthwhile.

Acknowledgments

This work was supported by a Research Grant from HFSP (Ref.-No: RGEC33/2024; https://doi.org/10.52044/HFSP.RGEC332024.pc.gr.194170). M.R-G also acknowledges support from grant PID2023-147067NB-I00 and grant RYC2021-032055-I funded by MCIU/AEI/10.13039/501100011033 and by ERDF/EU.

References

Aggarwal, I., Woolley, A. W., Chabris, C. F., & Malone, T. W. (2019). The impact of cognitive style diversity on implicit learning in teams. *Frontiers in psychology*, *10*, 112.

Allen, K., Brändle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., Griffiths, T. L., Hartshorne, J. K., Hauser, T. U., Ho, M. K., et al. (2024). Using games to understand the mind. *Nature human behaviour*, *8*(6), 1035–1043.

Biswas, U., Palod, V., Bhambri, S., & Kambhampati, S. (2025). Who is helping whom? analyzing inter-dependencies to evaluate cooperation in human-ai teaming. *arXiv preprint arXiv:2502.06976*.

Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., & Schulz, E. (2023). Empowerment contributes to exploration behaviour in a creative video game. *Nature Human Behaviour*, *7*(9), 1481–1489.

Carroll, M., Shah, R., Ho, M. K., Griffiths, T., Seshia, S., Abbeel, P., & Dragan, A. (2019). On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems*, *32*.

Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (2014). Reading the mind in the eyes or reading between the lines? theory of mind predicts collective intelligence equally well online and face-to-face. *PloS one*, *9*(12), e115212.

Games, G. T. (2016). Overcooked. *Team17: Wakefield, UK*.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385–16389.

Janssens, M., Meslec, N., & Leenders, R. T. A. (2022). Collective intelligence in teams: Contextualizing collective intelligent behavior over time. *Frontiers in psychology*, *13*, 989572.

Kim, Y. J., Engel, D., Woolley, A. W., Lin, J. Y.-T., McArthur, N., & Malone, T. W. (2017). What makes a strong team? using collective intelligence to predict team performance in league of legends. *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, 2316–2329.

Knott, P., Carroll, M., Devlin, S., Ciosek, K., Hofmann, K., Dragan, A. D., & Shah, R. (2021, January). Evaluating the Robustness of Collaborative Agents.

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.10.5].

Li, Y., Zhang, S., Sun, J., Du, Y., Wen, Y., Wang, X., & Pan, W. (2023). Cooperative Open-ended Learning Framework for Zero-Shot Coordination. *Proceedings of the 40th International Conference on Machine Learning*, 20470–20484.

Malone, T. W., & Bernstein, M. (2015). *Handbook of collective intelligence*. MIT press.

Matignon, L., Laurent, G. J., & Fort-Piat, N. L. (2012). Independent reinforcement learners in cooperative Markov games: A survey regarding coordination problems. *The Knowledge Engineering Review*, *27*(1), 1–31.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences*, *118*(21), e2005737118.

Rocca, R., & Tylén, K. (2022). Cognitive diversity promotes collective creativity: An agent-based simulation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Straub, V. J., Tsvetkova, M., & Yasseri, T. (2023). The cost of coordination can exceed the benefit of collaboration in performing complex tasks. *Collective Intelligence*, *2*(2), 26339137231156912.

Strouse, D. J., McKee, K. R., Botvinick, M., Hughes, E., & Everett, R. (2022, January). Collaborating with Humans without Human Data.

Vélez, N., Christian, B., Hardy, M., Thompson, B. D., & Griffiths, T. L. (2023). How do humans overcome individual computational limitations by working together? *Cognitive science*, *47*(1), e13232.

Verhagen, R. S., Neerincx, M. A., & Tielman, M. L. (2022). The influence of interdependence and a transparent or explainable communication style on human-robot teamwork. *Frontiers in Robotics and AI*, *9*.

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, *24*(6), 420–424.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *science*, *330*(6004), 686–688.

Wu, S. A., Wang, R. E., Evans, J. A., Tenenbaum, J. B., Parkes, D. C., & Kleiman-Weiner, M. (2021). Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, *13*(2), 414–432.

Xiang, Y., Vélez, N., & Gershman, S. J. (2023). Collaborative decision making is grounded in representations of other people's competence and effort. *Journal of Experimental Psychology: General*, *152*(6), 1565.

Yu, C., Gao, J., Liu, W., Xu, B., Tang, H., Yang, J., Wang, Y., & Wu, Y. (2023, February 3). *Learning Zero-Shot Cooper-*

ation with Humans, Assuming Humans Are Biased. arXiv: 2302.01605 [cs].

Zhang, Y., Sreedharan, S., & Kambhampati, S. (2016). A Formal Analysis of Required Cooperation in Multi-Agent Planning. *Proceedings of the International Conference on Automated Planning and Scheduling*, *26*, 335–343.

Zhao, M., Simmons, R., & Admoni, H. (2022, November). Coordination with Humans via Strategy Matching.