Role Stability and Team Performance in a 4-Player Cooperative Cooking Game

Sounak Banerjee (baners8@rpi.edu)
Cognitive Science Department, Rensselaer Polytechnic Institute

Wayne D. Gray (grayw@rpi.edu)
Cognitive Science Department, Rensselaer Polytechnic Institute

Abstract

The Cooperative Action Task (CAT) is a platform for studying the development of team coordination in complex dynamic task environments. Teams of four cooperate to play a cooking video game across eight 1-hr sessions. Team members communicate using gaze cursors that display the gaze location of each player. Team coordination in the game is achieved through a combination of planned and adaptive actions. Planned actions involve players acting according to preassigned roles to reduce behavioral variability, while adaptive actions are characterized by dynamic adaptations to changing task demands. The results of the study reveal that strategic reduction of behavioral variability was beneficial to game performance for all teams. Additionally, team performance was lower when teams switched between strategies across games in the same kitchen.

Keywords: Action Coordination; Games; Teams; Complex Task; Coordination Strategy; Team Roles; Planning; Adaptation

Introduction

There has been a rising interest in team research among cognitive scientists due to its significance in virtually every form of human coordination. However, analysis of team behavior in complex tasks can be challenging due to the dynamic nature of human interactions. Simulated virtual environments are an excellent tool for such analyses because they offer the complexity of naturalistic tasks while ensuring sufficient control over the task environment (Elliott et al., 2017; Cooke, Rivera, Shope, & Caukwell, 1999).

Computer games are excellent simulations for studying complex human behavior, especially expert behavior and task learning (Gray, 2017). For example, Tetris based studies have shed light on the various advanced strategies that experts use in the game and their implications on human learning (Gray & Banerjee, 2021; Sibert, Gray, & Lindstedt, 2020; Lindstedt & Gray, 2013). Others focused on differences in cognitive abilities among novices and experts (Large et al., 2019; Green & Bavelier, 2003).

For the current study, we developed a cooperative cooking game called "The CAT"; that is, the Cooperative Action Task. Here, the CAT was used to explore the development of team coordination (in 4-player teams) across eight 1-hour gameplay sessions. The experimental setup was further designed to enforce restrictions on communications within the team: players were prevented from verbally communicating with other team members during gameplay. However, players were allowed to communicate through a gaze-based communication system.

Current literature on cooperative behavior in humans reveals that coordinating humans rely on strategic reductions in action variability to improve action predictability for partners when communication is limited. One such study explored behavior in coordinating dyads in an action synchronization task, where access to information about partner's actions was limited (Vesper, Schmitz, Sebanz, & Knoblich, 2013). The authors discovered that subjects reduced action variability and improved coordination by speeding up their movements.

In the current study, teams reduced behavioral variability of players by assigning roles to its members. The teams used this strategy to compensate for the lack of a rich communication channel. To test for player persistence in sticking to assigned roles and its effect on team performance: we define 'Role Stability' (RS)— a measure of a player's tendency to stick to a certain role for the duration of a game. Results show that reduction of behavioral variability through adherence to player roles did improve team performance.

Methodology

Experimental Setup

The setup for the experiment is illustrated in Figure 1. It includes 5 computers, 4 eye-trackers (each attached to a monitor), 4 Xbox controllers, and 4 acoustic pods. Each pod had one controller, one eye-tracker, and one monitor inside. All 5 computers were set up outside the pods.

One of the 5 computers was used to run the game (the central node), and the video output for this computer was mirrored across all 4 monitors using an HDMI splitter. This meant all four players simultaneously received the same video stream for the game inside each pod. Since the game ran on the central node, the telemetry data (player actions and game state information) was also locally stored there at 60 frames per second (60Hz). The remaining 4 computers (edge nodes) were each connected to one of the 4 eye-trackers and placed outside the pods. Each edge node collected gaze data from the connected eye-tracker, stored the data locally, and sent it to the central node over the Local Area Network.

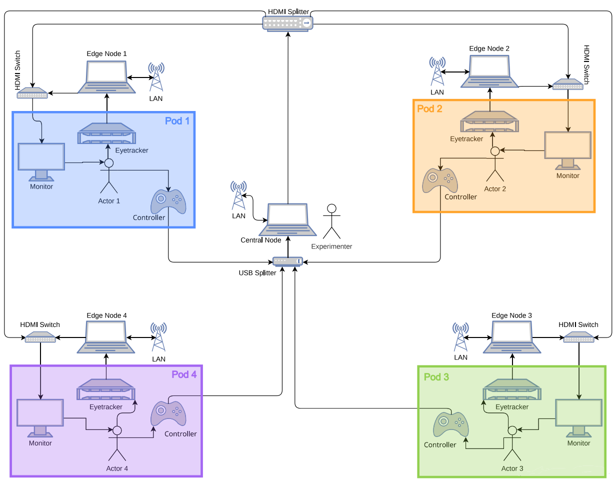


Figure 1: Layout of the experimental setup for the CAT. The 4 translucent regions correspond to the 4 pods in the setup. All entities enclosed within each one of the translucent regions represent the content of the pods.

The Cooperative Action Task

We present the Cooperative Action Task (The CAT), a game-based experimental paradigm (developed using the Unity game engine) to study human coordination within small 4-person teams in a controlled virtual environment. The goal of each team is to work together inside a virtual kitchen to prepare and deliver orders on time.

Orders appear at the top of the game interface, along with a timer indicating the time remaining to prepare the order. The example in Figure 2 presents two outstanding orders, a mushroom soup (expires in 35 seconds) and an onion soup (expires in 65 seconds). Players execute a series of actions to prepare each order as they come in. For example, to prepare the mushroom soup (from Figure 2) players from the team would have to chop three mushrooms and one onion (at the chopping counters), cook them in a pot (on a stove), plate the soup and carry it to the delivery zone. A dirty plate appears on the plate holder 10 seconds after each delivery. Players must then wash the dirty plate at the sink to prepare for the next order. Progress bars are used to indicate the progress of the cooking, chopping, and washing processes. Finally, if an item burns from being left on the stove too long, players need to dispose of it in the trash.

In the current version of the system, only gaze-based communication within teams was allowed during gameplay. To eliminate the possibility of any verbal communication, each player was placed in individual acoustic pods. Point-of-gaze was indicated using translucent disc-shaped gaze-cursors on the game interface, one corresponding to each player (see ‘ Gaze cursor’ labels in Figure 2). Every player could see all four gaze cursors on their screen, thus giving each team member access to others’ gaze locations. Further, players could also draw attention to their own gaze cursors by making their cursors pulse rapidly for half a second; this could be achieved by pressing a button on their controller.

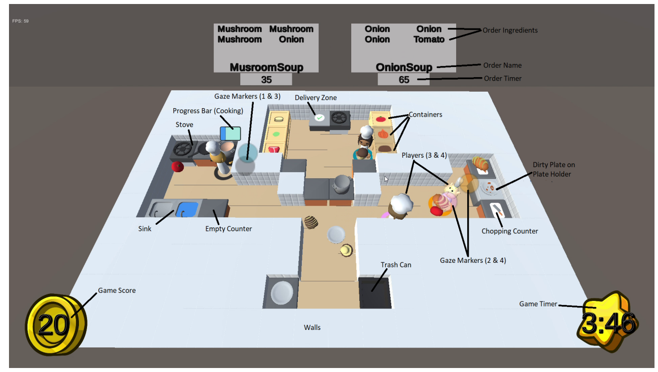


Figure 2: A (labelled) screenshot of a game in the ‘Clover’ kitchen layout. In this layout, the player at the top is locked out of the rest of the kitchen and is the only player with access to raw ingredients and the delivery zone.

Each game is a combination of a kitchen layout and an order list. Kitchen layouts are task environments that present unique challenges to team coordination, while order lists are used to tune the task’s difficulty by varying the amount of time available to prepare orders. Teams are awarded a score for each order they correctly deliver. The score for a specific order is 10 times the number of ingredients in the order. The theoretical maximum score possible for each order list is a function of the number and types of orders in the list.

During each session, teams played eight 5-minute games in 4 kitchen layouts (2 games per kitchen). Every pair of consecutive sessions shared the same set of 4 kitchen layouts. For example, games in sessions 1 and 2 were played in kitchen layouts 1 through 4, sessions 3 and 4 used layouts 5 through 8, and so on. So, each team played 4 games per kitchen. 16 kitchen layouts were used for the study; each layout presented a combination of various task constraints. Constraints included lack of space (counter space/floor space), narrow corridors, isolated players, and partitioned kitchens.

All 16 kitchen layouts were combined with 12 unique order lists to generate 64 unique games. The number of orders in any order list was kept high enough to ensure none of the teams would be able to complete all orders in the list.

Participants

The participants were 24 university students (9 female and 14 male and one participant chose not to answer). Participant age ranged from 19-27 years (mean=20.6, SD=1.84). All participants were between the ages of 19 and 22, except one 24- and one 27-year-old. Only one of the 6 teams was a homogeneous all-male team, the rest had both male and female members.

A campus-wide announcement was made for the study. 24 participants were selected from a pool of 40 students who expressed interest in the study. The selection criterion was based on the feasibility of all 4 participants being able to come into the lab (together) at least 3 times a week. Groups of 4 people with similar schedules were selected, it was done to minimize the number of canceled sessions due to the unavailability of one or more individuals. All experimental procedures were reviewed and approved by University IRB.

Procedure

Participants were first brought in for an introductory session, during which: (1) The study requirements and the participants' responsibilities were explained. (2) Subject IDs and team numbers were assigned, which remained constant for the entire duration of the study. (3) Three 1-hour timeslots were allotted to each group based on the availability of all 4 members. Two of the three timeslots were selected for the group's usual weekly schedule, that is, when they would come to the lab each week for the study. The third timeslot was used as a fallback option for rescheduling sessions, if necessary.

The study required participants to come to the lab for 11 one-hour sessions. The 11 sessions were executed in the following order: (1) In the first session, participants completed a Cognitive Task Battery (CTB) of 7 tasks; (2) the next 4 sessions (sessions 2-5), participants played the game; (3) during the sixth session, participants completed the Advanced Raven's Matrices test; (4) this was followed by 4 more game sessions (sessions 7-10); (5) in the final session, the CTB from session 1 was repeated.

Each game session involved participants playing eight 5-minute games (40 minutes total). Each player played the game inside their assigned (by the experimenter) acoustic pods. After playing the first 4 games, players were asked to step out of their pods and take a short break before returning to their pods to play the last 4 games of the session. Players were encouraged to discuss game strategy during the session breaks and at the end of each session. The experimenter on duty manually logged these discussions.

Data

The data analyzed in this study was obtained from six university student teams, each playing 64 games across the 8 game sessions. Data from one game was lost due to technical problems (Game 2 for Team 2). So, we performed this analysis using data from 383 games. Action, game state, and gaze information were recorded at 60Hz by the system. Game state information included the position of every object and player at every frame, score, active orders, time remaining per active order, and time remaining for the game. Action data included all button-press information for players and the resulting action within the game environment.

Analysis and Results

We began our analysis by plotting the average performance across all games played in each kitchen layout to test for the effects of various kitchen constraints on task performance. Figure 3 represents the average performance across all games in each kitchen layout. Six teams participated in the study, each playing 4 games per kitchen. This meant, we had data from 24 games for each kitchen layout, with the exception of the 'BaseLevel' kitchen, which had 23 data points because data from one game was lost due to technical issues.

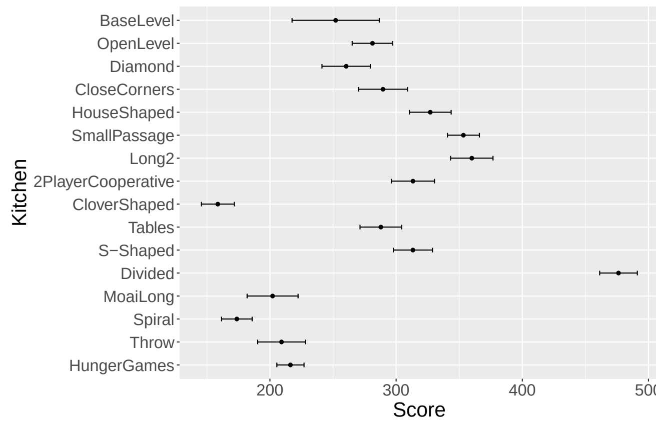


Figure 3: The graph presents the mean and standard error of scores across all games played in each kitchen. The kitchen layouts are arranged in chronological order (the sequence in which teams played games in the kitchens) from top to bottom.

Figure 3 presents several interesting trends for changes in performance across kitchens, which include increasing performance for games played in the first 7 kitchens and the relatively low performance in the final 4. However, in the current study we focus on the 'Divided' kitchen layout because of the consistent and considerably high scores associated with the games played in this kitchen. This was confirmed using a Tukey's HSD test which showed that the game scores for the Divided kitchen differed significantly ($p < .05$) from all other kitchens.

The high scores in the Divided kitchen are particularly intriguing because it is the only kitchen design where each of the four players was placed in separate sections of the kitchen and forced to work in isolation (Figure 4). Additionally, Teams played several games in 11 other kitchens before playing in the Divided kitchen. All 11 kitchen layouts were designed to allow (and, in some cases, force) players to collaborate with each other. Yet, none of the teams were able to adopt a cooperative strategy which was more efficient than working in isolation.

Interestingly, the Divided kitchen was not entirely devoid of coordination among team members. For example, to ensure multiple players did not end up preparing the same order, each individual had to keep track of the orders others were working on. Given the fast-paced nature and the complex structure of the game, in addition to the frequent overlap of ingredients in many orders, it was challenging to keep track of everything. However, based on experimenter observation, apart from a small number of instances, players were able to prepare orders without redundancy. Indeed, to reduce uncertainty, some teams used gaze cursors to indicate the orders on which they were working.

Figure 4: The 'Divided' kitchen layout. In this layout, every player is isolated to their own small kitchen with all necessary resources.

Teams also used designated player roles to reduce uncertainty during coordination on multiple occasions (teams discussed these strategies during session breaks and at the end of sessions). Pre-assigning player roles reduced each players' action variability, improving the team's predictability for player behavior, which ultimately aided coordination. Responsibilities for different player roles included chopping, cooking, and fetching (moving items around the kitchen for various purposes) items. Washing roles were almost never assigned because it is a relatively rare event and was always handled on the fly. So, washing actions were excluded from the analysis.

A correlation analysis of the different actions indicated that players who performed more cooking actions were also more likely to fetch items (0.31), while, chopping actions were negatively correlated with both cooking (-0.12) and fetching (-0.13). All correlations were statistically significant ($p < 0.05$). The correlations between different actions indicate players' tendency to organize their behavior around certain actions (roles) in the game.

To study the effects of player roles on team performance, we use 'Role Stability' (RS) to measure a player's tendency to adhere to specific roles in a game. We must first define 'Action Vectors' (AV) before we define role stability. An action vector is simply a 3-dimensional vector assigned to each player representing their contributions to different actions in a specific game. The 3 components of the vector represent the percentage of cooking, chopping, and fetching actions performed by each player. For example, the value of the cooking component of a particular player for a specific game is obtained using the following formula:

$$\frac{N^{P}_{cooking}}{N^{Total}_{cooking}} * 100$$

Where $N^{P}_{cooking}$ is the number of cooking actions executed by the player P in a game, and $N^{Total}_{cooking}$ is the total number of cooking actions executed by all players in that game.

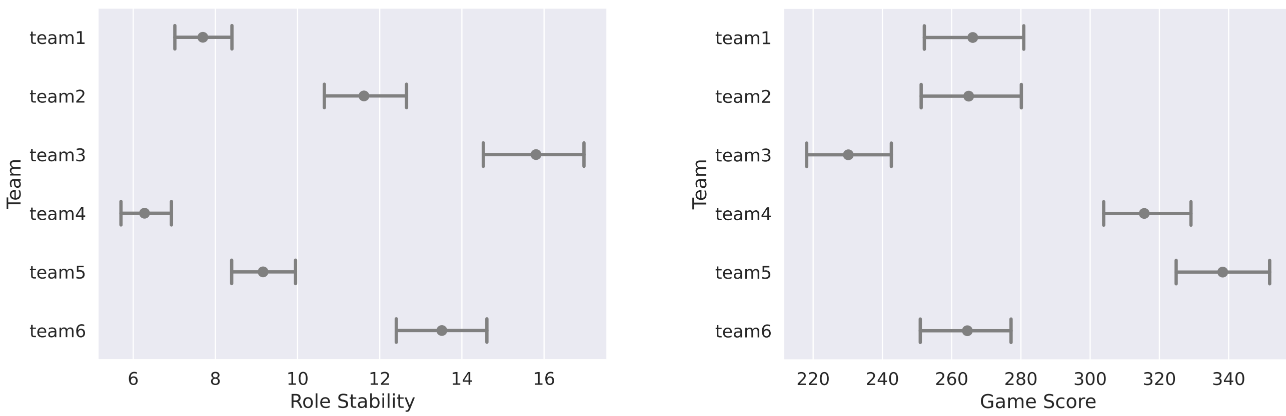
RS of a player in a specific game is simply the standard deviation of an AV. The value of RS is low when the values of the components for the corresponding AV are relatively similar (players engaging in all actions uniformly), while a high RS indicates one or two components have relatively higher values (players engaging more in specific activities). So, higher RS values indicate greater adherence to player roles, and lower RS values suggest more adaptation in players (switching roles as necessary).

Washing actions were excluded from AVs and, consequently, the measure of RS because the number of washing events in a game was negligible compared to other actions, and did not contribute sufficiently to the goals of the current analysis.

To study the relationship between RS and team performance, we first obtained a single measure of RS for each game that reflects the overall strategy used in the game by the team that played it; we refer to this as RS_{game}. RS_{game} is calculated by averaging the RS values of all 4 players in each game. A high value indicates a greater affinity among players to stick to existing roles, while a low value indicates a general adaptive behavior in the team. Figure 5 presents the mean and standard errors (grouped by teams) for RS_{game} values (5a) and game score (5b) for all 383 games.

Comparison of the two plots in Figure 5 reveals that the two highest scoring teams (Teams 4 and 5 in Figure 5b) had some of the lowest RS_{game} values (Figure 5a), while, the range of RS_{game} values for the worst performing team (Team 3) was the highest. Teams 2 and 6 demonstrated mediocre performance, and their RS_{game} values also hovered somewhere in the mid-ranges. Finally, the performance of Team 1 matched those of Teams 2 and 6 but their RS_{game} values were very low. The overall trend indicates an inverse relationship between RS_{game} and team performance (with a correlation of -0.28).

To account for the hierarchical nature of the data, mixed effects regression models were used to further test for the effect of role stability on team performance. However, in addition to RS_{game}, we define a second composite variable for role stability to use in the model. RS_{game} conveys information about the strategy used by a team in a specific game but fails to capture team behavior across games in a task environment (kitchen layout). The new variable was used in the model to add information about the spread of RS values across games played by a team in a kitchen. The values for the new variable were obtained by calculating the standard deviation of RS_{game} values for all 4 games played by each team in each kitchen layout. A high value of the standard deviation indicates greater variations in team strategy for games played in a specific kitchen, while a low value suggests use of similar strategies across games in a specific kitchen layout. Since the new variable provided a measure of a team's tendency to stick to similar strategies in a specific kitchen layout, we call this variable strategy consistency (SC).



(a) Plot for distribution of RS_{game}

(b) Plot for distribution of game score

Figure 5: Mean and standard error plots for RS_{game} (left) and game score (right) across all games played by each team.

Random Effect		
Variable	Variance	Std. Dev.
Team	1976	44.5
Kitchen Layout	7288	84.9

Fixed Effects			
Variable	Coeff.	Std. Err.	t value
Intercept	123.1	54.72	2.25
RS_{game}	4.24	1.19	3.55
SC	-4.21	3.74	-1.26
Baseline	0.34	0.12	2.79

Table 1: Results of model fit for a mixed effects model predicting team performance based on the value of RS_{game}, SC, and baseline performance for each team. The model also has random intercepts for each team and each kitchen.

The game score of each game was used as the dependent variable to fit mixed effects models. Random intercepts were used for Team IDs and kitchen layouts. RS_{game} and SC values were used as the fixed effect predictor variables. In addition, a baseline score for each team was also added as a fixed effect. The baseline score for a team was set to the score of the first game of the second session. Games from session 1 were not used for this purpose because teams spent their first sessions familiarizing themselves with the game mechanics. So, by using a game from the first session we open ourselves up to the possibility of selecting a game that might not be an accurate measure of the team's baseline performance.

Three models were fit to the data to determine the usefulness of RS_{game} and SC in predicting game score. The first model was a simple random effects model with random intercepts for each team along with the baseline performance of teams as a fixed effect (null model). For the second model, RS_{game} was added as a fixed effect to the null model. For the third and final model, SC was added as fixed effect to the second model. Model fits were assessed using the Akaike Information Criterion (AIC).

The model with no fixed effects was the worst of the three models (AIC: 4510). The model with only RS_{game} as a fixed effect was an improvement over the first model (AIC: 4498). Finally, the model with both fixed effects was the best model (AIC: 4493). Changes in AIC values between the 3 models were statistically significant, which indicate that both measures added predictive power to the model. The results of the final model fit are shown in Table 1. The positive coefficient for RS_{game} suggests that teams scored higher in games where they were more persistent about sticking to their roles compared to games in which they showed more adaptive behavior, while the negative coefficient for SC indicates that teams that were more likely to stick to a strategy across games for a particular kitchen design, performed better in general.

Discussion

We used a cooperative cooking game (The CAT) to study human coordination in a complex dynamic task. Six 4-player teams of university students each played the game across eight 1-hour sessions. Team communication was limited to a shared-gaze paradigm implemented in the system. However, teams were allowed to discuss gameplay outcomes and strategies during session breaks and at the end of each session.

Our data suggests that all 6 teams reached peak performance when players were isolated in their own sub-kitchens and forced to work alone (in the 'Divided' kitchen layout). In this layout, each player had to perform all actions necessary to prepare an order, including chopping, cooking, and fetching items. In other kitchen designs, where the kitchen is shared between players, teams would share responsibilities among players. However, effectively dividing responsibilities during gameplay was challenging in the absence of verbal communication channels. So, players often stuck to pre-assigned roles (decided by the team) to reduce prediction uncertainty and improve coordination within the team.

We use 'Role Stability' to measure players' tendency to stick to specific roles in a game. Due to the hierarchical structure of the data, we fit mixed effects models (with random intercepts for teams and kitchens) to determine the relationship between role stability and game performance. The results of the analysis suggest a positive relationship between role stability and performance, that is, reduction in behavioral variability through adherence to assigned player roles was beneficial to performance in general.

The results of the hierarchical model seemingly contradict the patterns observed in figure 5, which indicates an inverse relationship between role stability and game performance. However, this is not the case, as the apparent differences between the two results may be attributed to the random effects of team behavior and kitchen designs. Indeed, the plots in figure 5 indicate that teams which used more adaptive strategies on average (lower values of mean role stability) had higher game scores overall. Higher performance among adaptive teams in dynamic tasks have been suggested in the past. In one such study, teams that showed higher adaptability to role structure changes performed better when they were faced with unforeseen changes in the task (LePine, 2003). Dynamic allocation of team roles have also been shown to benefit team performance among artificial agents playing video games (Kim, 2006). Future publications may consider a deeper analysis of the inter- and intra-team differences in coordination.

Finally, The results also show that teams that stuck to similar strategies across games in each kitchen design were more likely to score higher. Teams that frequently switched strategies may have been experimenting with different strategies to find one optimal for the team, which could have led to poor performance.

Conclusion

We introduced – The CAT – a game-based experimental paradigm for studying team coordination. Each team played a cooperative cooking game across eight 1-hour sessions. Data was collected from six 4-player teams.

Team coordination strategies in the CAT belong to a spectrum between fully adaptive and fully planned behavior. Participating teams used a combination of strategies which involve assignment of player roles (planned behavior) and dynamic adaptation to changing task demands (adaptive behavior), for coordination. Assignment of player roles reduced behavioral variability thus increasing action predictability among team members, which in turn helped with team coordination. Our results show that strategic reduction of behavioral variability through adherence to player roles was beneficial to team performance. Finally, the results also show that teams performed worse when there was higher variation in their gameplay strategies across games.

References

Cooke, N. J., Rivera, K., Shope, S. M., & Caukwell, S. (1999, sep). A Synthetic Task Environment for Team Cognition Research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(3), 303–308. doi: 10.1177/154193129904300337

Elliott, L. R., Coovert, M. D., Schiflett, S. G., Salas, E., Cooke, N. J., & Shope, S. M. (2017). Designing a Synthetic Task Environment. In *Scaled worlds: Development, validation and applications* (1st ed., p. 273–288). Routledge.

Gray, W. D. (2017, mar). Game-XP: Action Games as Experimental Paradigms for Cognitive Science. *Topics in Cognitive Science*, *9*(2), 289–307. doi: 10.1111/tops.12260

Gray, W. D., & Banerjee, S. (2021, oct). Constructing Expertise: Surmounting Performance Plateaus by Tasks, by Tools, and by Techniques. *Topics in Cognitive Science*, *13*(4), 610–665. doi: 10.1111/tops.12575

Green, C. S., & Bavelier, D. (2003, may). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. doi: 10.1038/nature01647

Kim, I.-C. (2006). Dynamic role assignment for multi-agent cooperation. In A. Levi, E. Savaş, H. Yenigün, S. Balcısoy, & Y. Saygın (Eds.), *Computer and Information Sciences – ISCIS 2006* (pp. 221–229). Berlin, Heidelberg: Springer Berlin Heidelberg.

Large, A. M., Bediou, B., Cekic, S., Hart, Y., Bavelier, D., & Green, C. S. (2019, December). Cognitive and Behavioral Correlates of Achievement in a Complex Multi-Player Video Game. *Media and Communication*, *7*(4), 198–212. doi: 10.17645/mac.v7i4.2314

LePine, J. A. (2003). Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, *88*, 27-39. doi: 10.1037/0021-9010.88.1.27

Lindstedt, J. K., & Gray, W. D. (2013). Extreme Expertise: Exploring Expert Behavior in Tetris. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society, cogsci 2013, berlin, germany, july 31 - august 3, 2013.* cognitivesciencesociety.org.

Sibert, C., Gray, W., & Lindstedt, J. (2020). Interrogating Feature Learning Models to Discover Insights Into the Development of Human Expertise in a Real-Time, Dynamic Decision-Making Task. *Topics in Cognitive Science*, *9*, 374-394. doi: 10.1111/tops.12225

Vesper, C., Schmitz, L., Sebanz, N., & Knoblich, G. (2013). Joint Action Coordination through Strategic Reduction of Variability. *Annual Meeting of the Cognitive Science Society (cogsci)*(October), 1522–1527.