Toward a Framework for Characterizing Teaming Difficulties in Cooperative Video Games

Ziyu Li, Erik Harpstead

Human-Computer Interaction Institute
Carnegie Mellon University, Pittsburgh, PA, 15213 USA
ziyul@andrew.cmu.edu, eharpste@andrew.cmu.edu

Abstract

Current HAT research lacks a consistent framework for analyzing task environments. Drawing from cooperative game studies and team cognition, we propose a preliminary framework defining seven independent and two dependent dimensions to structure HAT task analysis, providing a foundation for future empirical studies and structured experimentation in HAT.

Background and Motivation

A key focus in current human-agent teaming (HAT) research is understanding the collaborative dynamics between humans and AI in teams. In both game studies and team science literature, the task environment has been recognized as a major factor influencing team dynamics (Beznosyk et al. 2012; Harris and Hancock 2019). However, there is a lack of consistent language for discussing the task environment in the context of HAT. (O'Neill et al. 2022). In addition, previous HAT experiments have reported mixed performance outcomes, with both underperformance and overperformance of the human-agent team attributed to a variety of disparate factors that are difficult to unify under a common theoretical framework (Demir, McNeese, and Cooke 2017; Wang et al. 2016).

One relevant body of literature that examines the relationship between task environment and team dynamics is game studies. By analyzing the design space of cooperative games, we can identify key dimensions of task design that shape team and task dynamics. This approach offers a structured language for future experimentation and the systematic development of HAT.

Our Framework

We propose a preliminary framework designed as a structured analysis tool for cooperative task environments. Drawing on insights from cooperative game studies and HAT research, we identify seven dimensions that serve as independent variables in any task environment. These dimensions integrate well-established cooperative game design patterns (Seif El-Nasr et al. 2010; Farah, Dorneich, and Gilbert 2022) and critical task environment features that necessitate different teaming modalities to effectively leverage the complementary strengths of humans and AI (Metcalfe et al. 2021). Additionally, we take inspiration from team cognition literature, incorporating task-work and team-work cognitive load as estimates of the difficulty that team members face in maintaining a shared understanding of both task execution and collaborative coordination (Schelble et al. 2022). These two dimensions serve as dependent variables in our task environment analysis.

Independent Dimensions

State Space Estimate of the number of parameters required to fully represent the state of the game at any given moment. In some games, this number varies (typically increases) as the game progresses from early stages to later levels. However, we consider only the latter scenario, as it better reflects the capacity of the task environment. For example, *Sid Meier's Civilization® VI* has a vast state space, where each of the thousands of tiles on the map requires tens of parameters. In contrast, *Overcooked!* has a relatively small state space, where players mainly track the positions of teammates, resource points, and the order queue.

Decision Space Approximates the size of the set of available courses of action given any game state. This might seem difficult to estimate, as the virtual reality of most games is in a continuous space, thus leading to an infinite continuous decision space, but we are mainly interested in the size of the discrete decision space where each decision would lead to a meaningfully different result. For instance, *Command & Conquer: Red Alert* has a significantly larger decision space compared to *Overcooked!*.

Entropy Estimates how much of the game state and decision space change when they do change, i.e. how much new information do players need to take into account each time the task environment change. For example, massive multiplayer online shooter games like *Battlefield V* have a relatively higher entropy than simple proceduralized games like *Don't Starve Together*.

Information Certainty Assesses the degree of certainty regarding the information required to make decisions, whether about the current game state or long-term strategy. Games with low information certainty utilizes randomness or limited visibility (e.g., fog of war), restricting players to only the information within their visibility range, as seen in *For the King* and *Counter-Strike*. Conversely, games with high information certainty feature deterministic environments with near perfect information, such as *Tools Up!* and *Satisfactory*.

Urgency Estimates the speed at which team members must make decisions and respond to potential mishaps. On the low end, turn-based games with no time constraints, such as *Baldur's Gate 3*, allow for deliberate decision-making. On the high end, real-time esports games, such as *Dota 2*, demand continuous player input and rapid reactions.

Role Heterogeneity Assesses the degree to which players on the same team possess different abilities to perceive and act within the game world. For example, in *Human Fall Flat*, all players have identical capabilities, whereas in *League of Legends*, each team member has a distinct set of skills and base attributes. Interestingly, in games like *League of Legends*, which have established socially agreed-upon playstyles, heterogeneous characters are often categorized into metagame-defined subroles. Thus, while individual characters may have diverse abilities, they still conform to a structured set of roles.

Role Coupling Denotes the level of interdependence between the agency of team members. In loosely coupled games like *Sid Meier's Civilization® VI*, tasks do not require tight collaboration between players and allow for more independent performance. In closely coupled games such as *Pico Park*, the actions of a player are directly influenced by or depend on others, and vice versa (Beznosyk et al. 2012).

Dependent Dimensions

Task-work cognitive load Estimates how much mental processing the player needs to do to form a "shared task model" on how to approach a task, and that is synchronized between team members. This shared task model might include: task procedures, task strategies, likely contingencies, and environmental constraints (Mathieu et al. 2000).

Team-work cognitive load Assesses how difficult it is for members of the team to develop a "shared team mode" that encodes team-specific knowledge such as interdependencies or synergies of roles, information flow, understanding of teammate's skills and tendencies (Mathieu et al. 2000).

Future Works

Now that we have a preliminary taxonomy of the HAT task space, we can begin to formulate hypotheses and experiment with them in a more structured way. For example:

- Is role coupling negatively related to task-work cognitive load due to a clearer mapping between player agency and task delegation?
- Are major contributing factors to task vs. team work cognitive load categorically different?

Or we can ask broader questions related to HAT like:

- Do HAT researchers and cooperative game designers agree on rating games across dimensions?
- Do tasks where human-AI teams can outperform pure human or pure AI teams always require high role heterogeneity?

Furthermore, we aim to quantitatively measure these relationships and assess their impact on the empirical performance of different configurations of HAT teams. To do so, we have developed our in-house benchmark game *Vertical Farm*, which afford researchers to ask comparative questions by intentionally designing tasks that can stake out points along our proposed dimensions. This task environment, with a planned release date of late May, focus on challenges that emphasis the task and team dynamic rather than focusing on individual human or AI difficulty, and we believe this shift of perspective from agent developer to task/game designer will yield us a HAT testbed that is future proof.

Acknowledgements

This work was funded by Army Research Lab STRONG Program Award #W911NF2120101. The opinions expressed are those of the authors and do not represent the views of the sponsoring agency.

References

Beznosyk, A.; Quax, P.; Lamotte, W.; and Coninx, K. 2012. The effect of closely-coupled interaction on player experience in casual games. In *Entertainment Computing-ICEC 2012: 11th International Conference, ICEC 2012, Bremen, Germany, September 26-29, 2012. Proceedings 11*, 243–255. Springer.

Demir, M.; McNeese, N. J.; and Cooke, N. J. 2017. Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, 46: 3–12.

Farah, Y. A.; Dorneich, M. C.; and Gilbert, S. B. 2022. Evaluating Team Metrics in Cooperative Video Games. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 66, 70–74. SAGE Publications Sage CA: Los Angeles, CA.

Harris, J.; and Hancock, M. 2019. To asymmetry and beyond! Improving social connectedness by increasing designed interdependence in cooperative play. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–12.

Mathieu, J. E.; Heffner, T. S.; Goodwin, G. F.; Salas, E.; and Cannon-Bowers, J. A. 2000. The influence of shared mental models on team process and performance. *Journal of applied psychology*, 85(2): 273.

Metcalfe, J. S.; Perelman, B. S.; Boothe, D. L.; and Mcdowell, K. 2021. Systemic oversimplification limits the potential for human-AI partnership. *IEEE Access*, 9: 70242–70260.

O'Neill, T.; McNeese, N.; Barron, A.; and Schelble, B. 2022. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, 64(5): 904–938.

Schelble, B. G.; Flathmann, C.; McNeese, N. J.; Freeman, G.; and Mallick, R. 2022. Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP): 1–29.

Seif El-Nasr, M.; Aghabeigi, B.; Milam, D.; Erfani, M.; Lameman, B.; Maygoli, H.; and Mah, S. 2010. Understanding and evaluating cooperative games. In *Proceedings of the*

SIGCHI conference on human factors in computing systems, 253–262.

Wang, D.; Khosla, A.; Gargeya, R.; Irshad, H.; and Beck, A. H. 2016. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.