Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026378383

AI Teammate Behavior and Player Experience in Cooperative Board Games: A Review

Yufei Dong*

School of Computer Science, Sichuan University, Chengdu, China

**Corresponding author: Yufei Dong.*

Abstract

Cooperative board games combine strategic problem-solving with communication, shared decision-making, and social interaction. AI teammates can lower barriers for novice players, support solo play, and provide a controlled setting for studying human-AI collaboration. However, high task performance alone does not guarantee a positive player experience: an effective but opaque agent may reduce players' sense of agency, whereas a socially expressive agent with weak task competence may undermine trust. This review synthesizes research from game AI, human-computer interaction, and multi-agent systems to examine three behavioral dimensions of AI teammates: ability alignment, transparency of intent, and socio-emotional expression. It further discusses trust, perceived control, and social presence as potential psychological mechanisms linking AI behavior to enjoyment, collaboration quality, and continued engagement. The reviewed evidence suggests that successful AI teammates should balance competence, interpretability, adaptability, and social expressiveness rather than maximize win rate alone. Based on this synthesis, the paper proposes an "AI behavior–psychological mechanisms–player experience" framework, summarizes relevant implementation and evaluation methods, and identifies research priorities for more reliable and player-centered AI teammate design.

Keywords

cooperative board games, AI teammates, player experience, human-AI collaboration, behavioral design, literature review

1. Introduction

Cooperative games have become an important design category because they combine strategic challenge with communication, coordination, and collective decision-making. Cooperative board games are especially valuable for research: their rules and action spaces are usually well defined, yet successful play often depends on interpreting partners' intentions, negotiating plans, and maintaining a shared understanding of the game state. These characteristics make them useful testbeds for studying human-AI collaboration.

AI integration may serve several practical purposes. An AI teammate can enable solo play, guide inexperienced players, fill an incomplete team, or provide adaptive assistance during complex tasks. Recent studies have therefore used games such as Hanabi, Codenames, and Terra Mystica to investigate reinforcement

learning, ad hoc teamwork, language-based assistance, and multi-agent coordination. The role of AI has expanded from rule-based non-player characters to learning agents and large language model (LLM)-based assistants capable of natural-language interaction.

Despite rapid technical progress, players' evaluations of AI teammates remain inconsistent. The same agent may be described as helpful and coordinated by some players but confusing, intrusive, or frustrating by others. This divergence indicates that objective performance metrics, such as score or win rate, capture only part of the experience. The effects of AI behavior are likely mediated by whether players understand the agent, feel able to influence the joint task, and perceive the agent as a credible social partner.

Existing studies are distributed across game AI, human-computer interaction (HCI), and multi-agent systems, and they often emphasize different outcomes. Algorithmic work tends to prioritize task performance and generalization, whereas player-centered work focuses on trust, enjoyment, and perceived agency. Three questions are therefore central to this review: (1) Which behavioral dimensions of AI teammates most strongly shape player experience? (2) Through which psychological mechanisms do these behaviors influence players? (3) Which computational and evaluation methods are most suitable for implementing and studying these behaviors?

This paper conducts a focused narrative review of representative studies on cooperative game AI, ad hoc teamwork, trust-aware interaction, and LLM-supported collaboration. The goal is not to provide a statistically exhaustive meta-analysis, but to integrate findings that illuminate the relationship between teammate behavior and player experience. The review first identifies three behavioral dimensions, then examines implementation techniques and evaluation methods, and finally proposes a unified conceptual framework and future research agenda.

2. Key Dimensions of Teammate Behavior and Their Impact on Player Experience

The influence of AI teammates can be organized around three interrelated dimensions: ability alignment, transparency of intent, and socio-emotional expression. These dimensions correspond to three practical player concerns: whether the AI can contribute effectively, whether its behavior can be understood and anticipated, and whether interaction with it feels socially meaningful. Their relative importance varies with the communication structure and strategic demands of each game.

2.1 Ability Alignment: The Goldilocks Problem in AI Performance

Ability alignment concerns whether an AI teammate's competence and level of assistance match the player's needs. Perez adapted an AlphaZero-style self-play architecture to Terra Mystica, using a 9×26 board representation with 206 feature channels and an action space of 2,143 moves [1]. Although the study primarily evaluated algorithmic feasibility, it illustrates a broader design problem: an AI that is too weak increases the player's workload, while an AI that is overwhelmingly strong may dominate decisions and reduce the player's contribution.

The objective should therefore be calibrated competence rather than maximum performance. Appropriate calibration may involve adjusting search depth, risk preference, explanation frequency, or the extent to which the AI proposes rather than executes actions. For novice players, the agent may provide more explicit guidance; for experienced players, it may adopt a less directive role. Ability alignment is thus closely linked to perceived control and challenge balance, both of which can influence enjoyment and willingness to continue playing.

2.2 Transparency of Intent: Predictability and Explainability

Transparency of intent refers to the extent to which players can predict an AI teammate's actions and understand the reasons behind them. Predictability and explainability are related but distinct. Predictability concerns behavioral consistency across similar situations, whereas explainability concerns whether the player can reconstruct or receive a meaningful account of the agent's decision.

Research on Hanabi demonstrates the value of strategic predictability. Hu et al. developed an adaptive meta-agent that generated diverse partner models and updated beliefs about teammate strategies during play. The adaptive agent performed better than a generalist baseline when paired with partners drawn from related

strategy distributions [2]. For player experience, the relevant implication is that an agent that recognizes and responds consistently to a partner's conventions is easier to coordinate with than one whose behavior changes without an interpretable pattern.

Predictability alone, however, does not ensure understanding. In a Codenames study, Sidji et al. found that an LLM assistant could support idea generation while also complicating the formation of a shared team mental model [8]. Players sometimes struggled to distinguish human contributions from AI-generated ones and to determine how much trust to place in each source. Although the system functioned as an assistant rather than a fully independent teammate, the study highlights a general design requirement: AI contributions should be attributable, and explanations should clarify the agent's intention without interrupting the pace of play.

2.3 Socio-Emotional Expression: From Tool to Teammate

Socio-emotional expression concerns whether the AI is experienced as a responsive partner rather than merely an optimization tool. Relevant behaviors include acknowledging a partner's action, expressing uncertainty, requesting clarification, celebrating joint success, and adapting tone to the situation. These cues may increase social presence, but they are beneficial only when they remain consistent with the agent's actual capabilities and do not create misleading impressions of understanding.

Palatsi et al. provide indirect evidence that LLMs can reproduce patterns resembling human cooperation in classic game-theoretic settings. Across 121 two-player scenarios, Llama's decision distribution correlated strongly with human choices ($r = 0.84$) and more closely resembled human behavior than equilibrium predictions [3]. The study does not directly establish that socially expressive AI improves board-game experience, but it suggests that language models can represent trade-offs between cooperation and defection in ways that may support context-sensitive teammate behavior.

This technical potential should be treated cautiously. Human-like language can increase engagement, yet it may also encourage players to overestimate the system's understanding or reliability. Effective socio-emotional design therefore requires calibrated expression: the AI should communicate confidence and uncertainty accurately, avoid manipulative emotional cues, and use social behaviors to support coordination rather than distract from the shared task.

2.4 Interactions Among the Three Dimensions and Variation Across Game Types

The three dimensions interact rather than operate independently. Basic competence is necessary for useful collaboration, but competence without transparency may appear arbitrary or controlling. Transparency supports trust calibration by helping players understand what the AI can and cannot do. Socio-emotional expression can strengthen social presence and engagement, but only when grounded in reliable task behavior. Their combined influence can be represented as a pathway from AI behavior to psychological mechanisms—particularly trust, perceived control, and social presence—and then to outcomes such as enjoyment, team performance, and continued engagement.

Different games assign different weights to these dimensions. Terra Mystica emphasizes strategic competence and information processing. Hanabi depends heavily on consistent conventions and interpretable clues, making transparency especially important. Codenames relies on creative association, shared context, and interpersonal familiarity, so socio-emotional and communicative fit may matter as much as task efficiency. The same AI design is therefore unlikely to be optimal across all cooperative games.

Design priorities should be derived from the target game's collaboration structure. A useful development process begins by identifying what information players must share, where uncertainty arises, which decisions should remain under human control, and what forms of social interaction contribute to the game's appeal.

3. Computational Techniques for Implementing and Shaping AI Teammate Behavior

The behavioral dimensions above can be supported by several computational approaches. There is no one-to-one mapping between a technique and a behavioral outcome: adaptive learning may improve ability alignment and predictability, while symbolic constraints may improve transparency and safety. The following approaches illustrate the main design options and their trade-offs.

3.1 Self-Play and Search-Based Learning

Perez's Terra Mystica system is better characterized as self-play and search-based learning than as imitation learning. The model combined policy and value networks with Monte Carlo tree search and learned through repeated self-play rather than from human demonstrations [1]. Such methods can produce strong strategic behavior in games with large state spaces, but strength alone does not ensure compatibility with human partners. To support player experience, self-play agents may need additional objectives for human-likeness, diversity, or controllable difficulty, as well as interfaces that expose plans and uncertainty.

3.2 Adaptive Multi-Agent Learning

In dynamic collaborative settings, an AI must often coordinate with partners whose strategies were not represented explicitly during training. Ma et al. proposed PACE (Peer Adaptation with Context-aware Exploration), which operates in partially observable stochastic games. A context encoder models interaction trajectories, while a peer-identification objective encourages the agent to gather information about partner preferences. Across cooperative, competitive, and mixed-motive tasks, PACE distinguished among partner behaviors and adapted rapidly, particularly in cooperative scenarios [4].

The method addresses a central problem in ad hoc teamwork: identifying a partner's strategy from limited interaction. For board games, this capability could allow an AI to adjust its level of assistance or follow a player's preferred convention. However, adaptation can reduce predictability if behavioral changes are not communicated. Adaptive systems should therefore signal when and why they change strategy. Their performance may also decline when test-time partners differ substantially from the training population.

3.3 Data Augmentation Strategies

Data augmentation can improve generalization across strategic conventions. Hammond et al. proposed Symmetry-Breaking Augmentations (SBA), based on the observation that cooperative strategies often depend on arbitrary but internally consistent mappings, such as color conventions. SBA applies symmetry transformations to observations and actions during training, exposing the agent to multiple equivalent conventions [5].

In Hanabi, SBA-trained agents collaborated more effectively with previously unseen partners than baseline agents, especially when training data were limited. The authors also introduced the Augmentation Impact metric to estimate how strongly a transformation changes strategy diversity. SBA offers a relatively low-cost way to broaden compatibility, but it is applicable only when meaningful symmetries can be identified. It should therefore complement, rather than replace, partner modelling and online adaptation.

3.4 Symbolic Behavioral Architectures

Symbolic architectures, including behavior trees and finite-state machines, remain useful when designers require explicit control, auditability, and stable behavior. Behavior trees organize actions hierarchically through conditions and prioritized branches, while finite-state machines specify transitions between predefined states. Their logic can be inspected directly, making failures easier to diagnose and explanations easier to generate.

The main advantage of symbolic systems is predictable behavior under known conditions. This is valuable for tutorials, rule enforcement, and safety-critical restrictions on learned agents. Their limitation is reduced flexibility in open-ended or strategically complex situations. A practical solution is a hybrid architecture in which a learned component proposes actions while a symbolic layer enforces rules, manages dialogue states, and prevents clearly undesirable behavior.

4. Research Methods and Tools for Evaluating AI Teammates and Player Experience

Evaluating AI teammates requires mixed methods because no single metric captures both collaborative performance and subjective experience. A rigorous study should combine self-report measures, behavioral telemetry, and controlled experiments, while also considering repeated interaction and individual differences.

Subjective measures can assess enjoyment, frustration, trust, perceived control, social presence, and willingness to play again. Established game-experience questionnaires may be combined with collaboration-specific scales. Kraus et al. developed a trust-aware user simulator and questionnaire using 3,696 human-AI exchanges from 308 dialogues, and modelled changes in trust during interaction with proactive agents [6]. This work illustrates how trust can be treated as a dynamic variable rather than a single post-session rating. In board-game studies, measurements should distinguish appropriate trust from overtrust and distrust.

Objective analysis can use game telemetry to measure task completion, score, error recovery, response time, resource exchange, and the distribution of decision-making between the player and AI. Gong et al.'s MINDAGENT framework introduced the Cuisine World benchmark and a Collaboration Score for evaluating multi-agent efficiency [7]. GPT-4 acted as a centralized scheduler for teams of two to four agents, and ablation results showed that environmental feedback was essential for avoiding repeated errors. Although the environment is not a board game, the study demonstrates the value of logging intermediate actions and feedback, rather than evaluating only the final result.

Controlled experiments are needed to establish causal links between design choices and experience. Participants can be randomly assigned to AI conditions that vary in competence, explanation style, adaptability, or social expression. A/B comparisons should report both objective outcomes and subjective responses, and should control for player expertise, prior familiarity with the game, and expectations about AI. Repeated-session designs are especially important because trust, convention formation, and annoyance may change over time. Qualitative interviews or think-aloud protocols can further reveal why players interpret the same behavior differently.

5. Conclusion

This review examined how AI teammate behavior may shape player experience in cooperative board games. The evidence supports three central dimensions: ability alignment, transparency of intent, and socio-emotional expression. These dimensions influence experience indirectly through mechanisms such as trust calibration, perceived control, and social presence. Their importance depends on the game's strategic complexity, information structure, and social goals.

At the implementation level, self-play and search can provide strategic competence; adaptive multi-agent learning supports coordination with unfamiliar partners; data augmentation improves robustness across conventions; and symbolic architectures provide controllability and interpretability. Hybrid designs are particularly promising because they can combine learned flexibility with explicit behavioral constraints. At the evaluation level, subjective scales, telemetry, controlled experiments, and repeated interaction studies should be used together.

The proposed “AI behavior–psychological mechanisms–player experience” framework offers a way to connect algorithmic choices with player-centered outcomes. Its main implication is that an AI teammate should not be evaluated only by how often it wins. A successful agent should contribute at an appropriate level, communicate its intentions clearly, preserve meaningful player agency, and express social cues in a calibrated manner.

Several limitations remain. The current evidence base is small and heterogeneous, and some studies reviewed here concern adjacent settings such as dialogue systems or general multi-agent benchmarks rather than cooperative board games directly. Socio-emotional effects are especially under-tested, and claims about causality require more controlled player studies. Future research should examine long-term trust formation, cultural and expertise differences, the effects of LLM-generated explanations, and methods for adapting difficulty without making behavior unpredictable. Progress in these areas would support AI teammates that are not only more capable, but also more understandable, appropriately social, and enjoyable to play with.

References

[1] Perez, L. (2021). Mastering Terra Mystica: Applying self-play to multi-agent cooperative board games. arXiv. https://arxiv.org/abs/2102.10540

[2] Hu, H., Lerer, A., Hu, B., Foerster, J., & Brown, N. (2022). Generating and adapting to diverse ad hoc partners in Hanabi. IEEE Transactions on Games. Advance online publication. https://www.x-mol.com/paper/1534679146241527808/t

[3] Palatsi, A. C., Gächter, S., Kleiman, T., & McElreath, R. (2025). Large language models replicate and predict human cooperation across experiments in game theory. arXiv. https://arxiv.org/abs/2511.04500

[4] Ma, Y., Lu, C., Lv, K., Hao, J., Liu, T.-Y., & Fu, J. (2023). Fast Peer Adaptation with Context-aware Exploration. Proceedings of the 40th International Conference on Machine Learning (ICML), 2023. https://proceedings.mlr.press/v235/ma24n.html

[5] Hammond, R., Bou-Ammar, H., & Howley, E. (2024). Symmetry-breaking augmentations for ad hoc teamwork. arXiv. https://arxiv.org/abs/2402.09984

[6] Kraus, M., Riekenbrauck, R., & Minker, W. (2023). Development of a trust-aware user simulator for statistical proactive dialog modeling in human-AI teams. In Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (pp. 38–43). ACM. https://doi.org/10.1145/3563359.3597403

[7] Gong, R., Li, J., Zhu, J., Ma, Y., Chen, X., Chen, S., Ma, H., & Xu, H. (2023). MindAgent: Emergent gaming interaction. arXiv. https://arxiv.org/abs/2309.09971

[8] Sidji, M., Smith, W., & Rogerson, M. J. (2024). Human-AI Collaboration in Cooperative Games: A Study of Playing Codenames with an LLM Assistant. Proceedings of the ACM on Human-Computer Interaction, 8(CHI PLAY), 1-25. https://doi.org/10.1145/3677081

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access