

Who is Helping Whom? Analyzing Inter-dependencies to Evaluate Cooperation in Human-AI Teaming

Upasana Biswas, Vardhan Palod, Siddhant Bhambri, Subbarao Kambhampati

School of Computing and AI, Arizona State University, Tempe, AZ, USA
 {ubiswas2,vpalod,sbhambr1,rao}@asu.edu

Abstract

State-of-the-art methods for Human-AI Teaming and Zero-shot Cooperation focus on task performance, as the sole evaluation metric while being agnostic to ‘how’ the two agents work with each other. Furthermore, subjective user studies only offer limited insight into the quality of cooperation existing within the team. Specifically, we are interested in understanding the cooperative behaviors arising within the team when trained agents are paired with humans - a problem that has been overlooked by the existing literature. To formally address this problem, we propose the concept of constructive interdependence - measuring how much agents rely on each other’s actions to achieve the shared goal - as a key metric for evaluating cooperation in human-agent teams. We measure interdependence in terms of action interactions in a STRIPS formalism, and define metrics that allow us to assess the degree of reliance between the agents’ actions. We pair state-of-the-art agents with learned human models as well as human participants in a user study for the popular Overcooked domain, and evaluate the task reward and teaming performance for these human-agent teams. While prior work has claimed that state-of-the-art agents exhibit cooperative behavior based on their high task rewards, our results reveal that these agents often fail to induce cooperation, as evidenced by consistently low interdependence across teams. Furthermore, our analysis reveals that teaming performance is not necessarily correlated with task reward, highlighting that task reward alone cannot reliably measure cooperation arising in a human-agent team.

Code —

<https://github.com/upasana27/coop-eval-user-study.git>

Extended version — <https://arxiv.org/abs/2502.06976>

Introduction

Achieving zero-shot cooperation (ZSC)—i.e., enabling agents to collaborate effectively with previously unseen partners—is a key challenge in cooperative AI. This capability is particularly important in human-agent teaming (HAT) where agents must interact with and adapt to a diverse range of human behaviors. Popular approaches for developing such agents often rely on the task reward as the primary signal for learning and evaluation (Lou et al. 2023; Sarkar, Shih, and Sadigh 2023; Yu et al. 2023; Strouse et al.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

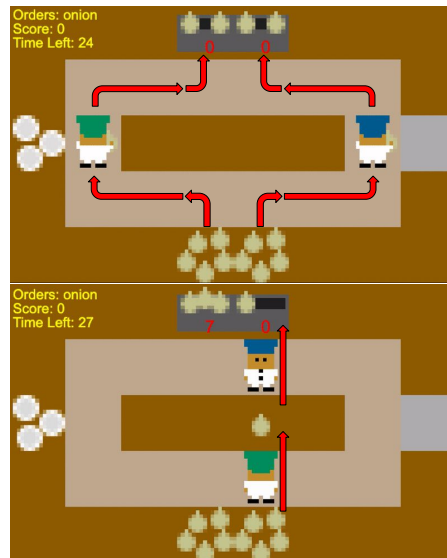


Figure 1: Depicted are two strategies to fill a pot with onions in a cooking game. The coordinated strategy (right) is more efficient than the individual strategies (left), but runs the risk of failure if cooperation is not achieved.

2022). However, task reward alone is often insufficient to fully capture collaborative behavior, as it can obscure important details about individual teammate performance and the interactions that arise between them. Agents may learn to optimize task success by operating independently, without adapting to their teammate’s actions. This issue is especially pronounced in domains where cooperation is not strictly required for task completion, as strong task performance may be attained with minimal or no coordination.

To illustrate this, we draw on the distinction between Required Cooperation (RC) and Non-Required Cooperation (Non-RC) settings introduced by Zhang, Sreedharan, and Kambhampati (2016). In RC scenarios, all team members must participate to achieve the shared goal, whereas in Non-RC settings, individuals can independently complete the task without relying on teammates. For example, consider a human-agent team playing Overcooked (Figure 1), where players must prepare and serve soups by collecting

and cooking onions. In this environment, collaboration is not strictly necessary—one player can complete the task while the other remains idle, or the two can engage in active coordination, such as passing onions. Crucially, both strategies can yield similar task rewards despite vastly different levels of cooperation and adaptation. This issue is especially concerning in human-agent teams involving ZSC agents, where humans may compensate for an idle or non-adaptive agent, resulting in high team task rewards that conceal deficiencies in the agent’s cooperative abilities.

The insufficiency of task reward to evaluate ZSC agents is rooted in the shadowed equilibrium problem (Matignon, Laurent, and Le Fort-Piat 2012; Fulda and Ventura 2007), which arises in Non-RC domains with multiple viable strategies to accomplish a task. In such settings, agents frequently converge to non-cooperative equilibria because cooperative strategies are rarely explored during training (Lerer and Peysakhovich 2019). As a result, ZSC agents may fail to recognize and reciprocate cooperative intentions when paired with partners who actively seek collaboration (Carroll et al. 2020). Evaluating agents solely through task reward conceals these failures of ZSC agents, which are especially detrimental in human-agent teams where unreciprocated cooperation can degrade trust, team cohesion, and overall effectiveness (Textor et al. 2022; Lopez et al. 2023). Consequently, relying solely on task reward limits the development and assessment of agents’ ability to adapt and effectively cooperate with diverse human teammates in real-world settings.

To assess cooperation arising in human-agent teams, we focus on **interdependence**—a structured form of cooperation where team members’ actions are contingent on one another (Johnson, Vignatti, and Duran 2020). This type of coordination is central to many real-world HAT domains such as Urban Search and Rescue (Pateria, Subagdja, and Tan 2019), collaborative trash collection (Ghavamzadeh, Mahadevan, and Makar 2006), and predator-prey systems (Wu et al. 2023; Barton et al. 2018). While prior work categorizes interdependence into pooled, sequential, reciprocal, and team-based (Verhagen, Neerincx, and Tielman 2022; Singh, Miller, and Sonenberg 2016, 2014), we specifically target sequential and reciprocal interdependencies. We propose a novel metric for measuring such interdependencies between a human and an agent working as a team. We map a two-player Markov game to a symbolic STRIPS-like formalism (Fikes and Nilsson 1971), introducing symbolic structure to both world states and actions.

We pair state-of-the-art ZSC agents with human partners and models (Wang et al. 2024a). We evaluate these teams using the proposed metric in the Overcooked domain—a popular benchmark where numerous approaches have been developed for zero-shot cooperation and HAT (Strouse et al. 2022; Carroll et al. 2020; Zhao et al. 2022; Yu et al. 2023; Li et al. 2024), thus making it an ideal testbed. The proposed metric generalizes to any domain that can be represented using a symbolic formalism. To support ease of integration of the proposed metric into other domains, we provide a fully modular software package (pseudocode in Technical Appendix A), and present a concrete application to a Search

and Rescue domain (Technical Appendix B). Our main contributions are:

- We propose a domain-agnostic metric that quantitatively captures sequential and reciprocal interdependence in human-agent teams.
- Using our proposed metric in the Overcooked domain, we demonstrate that task reward alone can be misleading as an indicator of cooperation in human-agent teams.
- We validate our metric with a user study and simulations involving state-of-the-art ZSC agents, revealing their limitations in adapting to cooperative partners.

Our analysis reveals a critical gap in current ZSC agents: when paired with partners actively seeking collaboration, these agents often fail to adapt. This issue is particularly pronounced in Non-RC settings, where teams may still achieve high task rewards despite minimal genuine coordination. These findings highlight the limitations of relying solely on task reward and emphasize the importance of explicit interdependence metrics to accurately assess cooperation in human-agent teams. We also present preliminary experiments showing that incorporating the proposed cooperation metric during training leads SOTA ZSC agents to consistently increase constructive interdependencies without harming task performance, suggesting that the metric can effectively promote cooperative behavior.

Related Works

Prior work in human-agent teaming often relies on task performance or episodic reward as evaluation metrics (Strouse et al. 2022; Yu et al. 2023; Zhao et al. 2022; Li et al. 2024; Wang et al. 2024a; Lou et al. 2023). Other approaches emphasize alternative metrics such as collaborative fluency, robot and human idle time (Zhao, Simmons, and Admoni 2022; Knott et al. 2021; Fontaine et al. 2021), or use subjective user studies to assess trust, engagement, and fluency (Zhao, Simmons, and Admoni 2022; Ma et al. 2022; Nalepka et al. 2021). However, these metrics are typically sensitive to specific environment layouts and task structures. Moreover, subjective evaluations offer limited insight into the underlying quality of cooperation. In contrast, our proposed interdependence metric provides a domain-general and objective measure of cooperative behavior between teammates.

Zhang et al. (2024) evaluate team outcomes and collaboration characteristics such as contribution rate, individual effort, and communication frequency. Bishop et al. (2020) propose action-based metrics like Productive Chef Actions (PCA), PCA duration, and Chef Role Contribution (CRC) to quantify individual roles during execution. Similarly, Ries et al. (2024) compute team contribution by comparing the proportion of tasks completed by humans and AI agents. While these approaches quantify task participation, they do not capture the underlying cooperative dynamics or structural dependencies between teammates’ actions.

Our interdependence metric complements recent work by Wang et al. (2025), which focuses on how human-agent teams adapt and evolve over time. While shared goals and team acceptance are often measured via subjective reports (Liang et al. 2019), our metric offers an objective sig-

nal of cooperative behavior by revealing when teammates act in ways that enable or anticipate each other’s contributions. Successfully formed and fulfilled interdependencies can indicate role adherence (Wang et al. 2024b), trust, and coordination (Moran et al. 2013; Cai et al. 2019).

Interdependence has also been positioned as a central organizing principle for human-machine teaming (Johnson et al. 2014; Johnson, Vignatti, and Duran 2020). Prior work has explored coordination through low-level dynamics using Convergent Cross Mapping (CCM) (Barton, Waytowich, and Asher 2018; Wu et al. 2023; Barton et al. 2018), whereas our approach targets symbolic, structured task dependencies. Verhagen, Neerinx, and Tielman (2022); Singh, Miller, and Sonenberg (2014) categorize interdependence into pooled, sequential, reciprocal, and team-based. We formalize interdependence when one agent’s action effect satisfies another’s precondition, using a STRIPS-based representation. This captures both unidirectional (sequential) and bidirectional (reciprocal) dependencies in human-agent collaboration.

Preliminaries

Two-Player Markov Game

A two-player Markov game for a human-AI cooperation scenario can be defined as $\langle S, A, T, R \rangle$ where S is the set of world states, $A : A_1 \times A_2$ where A_i is the set of possible actions for agent i , $T : S \times A_1 \times A_2 \rightarrow S$ is the transition function mapping the present state and the joint action of the agents to the next state of the world, and $R_i : S \times A_1 \times A_2 \rightarrow R_i$ is the reward function mapping the state of the world and the joint action to the global reward.

For a 2-player cooperative Markov game, $R = R_1 = R_2$ where R is the global environment reward function. The joint policy is defined as $\pi = (\pi_1, \pi_2)$ where the policy $\pi_i : S \rightarrow A_i$ is defined for an agent i over the set of possible actions A_i . The objective of each agent i is to maximize the expected discounted return $E_\pi [\sum_{t=0}^{\infty} \gamma^t R(s^t, a_1^t, a_2^t)]$ by following the policy π from a given state. Therefore, the policy π is learned by optimizing the task reward received by the agents from the environment.

Multi-Agent Planning Problem

A STRIPS (Fikes and Nilsson 1971) problem is represented as $\langle P, A, I, G \rangle$ where P is the set of propositions which can be used to denote facts about the world, A is the set of planning actions, I is the initial state, and G is the goal state. Each fluent $p \in P$ is a symbolic variable that describes the current state of the environment, with each proposition representing a specific property of an object in the world.

The possible fluents for the Overcooked environment can be *counter-empty* — describes whether the counter is empty or not, *pot-ready* — indicates whether the soup is ready in the pot, *soup-served* — indicates whether the soup has been served at the serving station, etc. I denotes the propositions representing the initial state of the world, and G denotes the propositions corresponding to the goal state of the world.

A planning action can be defined as $a = \langle \text{pre}(a), \text{add}(a), \text{del}(a) \rangle$ where $\text{pre}(a)$ is the set of proposi-

tions that must be true before the action can be executed, $\text{add}(a)$ are the propositions that become true after the action is performed, and $\text{del}(a)$ are the propositions that become false after the action is performed. Extending this to multiple agents, a Multi-Agent Planning task can be denoted as $\langle P, N, \{A_i\}_{i=1}^N, I, G \rangle$ where N is the number of agents and A_i is the set of actions for agent i .

We assume that agents act in parallel at each step of the plan, selecting and executing one action simultaneously. A plan is defined as a sequence of joint action sets $(\{a_i^1\}_{i=1}^N, \{a_i^2\}_{i=1}^N, \dots, \{a_i^n\}_{i=1}^N)$, where $\{a_i^t\}_{i=1}^N$ denotes the actions taken by all agents at timestep t , $\{a_i^t\}$ is the action taken by the i^{th} agent and n is the number of steps in the plan. A plan is a solution Π if it is a sequence of joint actions that can be applied to the initial state I and results in a world state that satisfies G , i.e., $\Pi = (\{a_i^1\}_{i=1}^N, \{a_i^2\}_{i=1}^N, \dots, \{a_i^n\}_{i=1}^N)$ is a valid solution plan if $\{a_i^n\}_{i=1}^N (\dots (\{a_i^2\}_{i=1}^N (\{a_i^1\}_{i=1}^N (I)))) \subseteq G$.

Interdependencies

Problem Statement

We pose the human-agent teaming problem as a two-player Markov game, where the teammates act in parallel. We focus on the case where the team is trying to reach a set of goal states S_G such that $S_G \subseteq S$. The states in S_G are absorbing i.e. $\forall s \in S_G$ and $a_i^G \in A_i$, we have $T(s, \{a_i^G\}_{i=1}^2) = 0$.

We represent the solution trajectory for the i^{th} agent as $\tau_i = (a_i^t, a_i^{t+1}, \dots, a_i^k \dots a_i^n)$ and the joint-action solution trajectory of two agents starting from timestep t and reaching a goal state at timestep n as $\tau = ((a_1^t, a_2^t), (a_1^{t+1}, a_2^{t+1}) \dots (a_1^n, a_2^n))$. An execution trace Tr of a policy π from an initial state s^t as is denoted as $(s^t, a^t, s^{t+1}, a^{t+1}, \dots, s^n)$, where Tr corresponds to the state-action sequence that starts at timestep t and terminates at a goal state $s^n \in S_G$ at a timestep n , where $a^k = (a_1^k, a_2^k)$ and $a_i^k = \pi_i(s^k)$ for the k^{th} timestep. The agents receive a task reward R_{task} at the end of Tr and τ on reaching the goal state.

Given the execution trace Tr , the joint solution trajectory τ and only the scalar task reward R_{task} , there is no explicit measure of the cooperation exhibited in τ . To capture the cooperative interactions arising between the teammates in τ , we define the concept of interdependence in the next section.

Mapping the Markov Game to STRIPS

In a Markov Game, the state at a given timestep $s_t \in S$ is typically represented as a high-dimensional vector. We can instead describe s_t as a symbolic state consisting of a set of true propositions p_t that denote relevant facts about the current world state. This allows us to represent each state as a finite set of meaningful symbolic facts.

Formally, there exists a function $\mathcal{F} : S \rightarrow 2^P$ that maps a state s_t to the corresponding set of true propositions p_t . For example, consider Fig. 1. The predicate *counter-empty* denotes whether the middle counter is empty. Suppose the green-hat agent (A_2) takes an action to place an onion on the counter. Before this action, the proposition *counter-empty* is

true in state s_t , but after the action, it becomes false in state s_{t+1} . Mapping states to symbolic propositions thus enables us to capture the effects of agents' actions in terms of relevant symbols.

Recall from the execution trace Tr of the Markov Game that at time t , the world is in state s^t . Taking joint action a^t causes a transition to the next state s^{t+1} . Using the mapping \mathcal{F} , we can represent this transition symbolically as (p_t, a^t, p_{t+1}) where $p_t = \mathcal{F}(s^t)$ and $p_{t+1} = \mathcal{F}(s^{t+1})$. Similarly, the joint action $a^t = (a_1^t, a_2^t)$ can be mapped to a symbolic representation. For each individual action a_i^t , there exists a corresponding STRIPS-style planning action such that $\text{pre}(a_i^t) \subseteq p_t$, $\text{add}(a_i^t) \subseteq p_{t+1}$ and $\text{del}(a_i^t) \subseteq P \setminus p_{t+1}$. Thus, the entire solution trajectory τ can be represented as a joint solution plan Π , where each single-agent action a_i^t is expressed as $a_i^t = \langle \text{pre}(a_i^t), \text{add}(a_i^t), \text{del}(a_i^t) \rangle$. This representation allows us to systematically track the preconditions and effects of individual agent actions in the trajectory using symbolic propositions, which in turn enables us to analyze and capture interdependencies between them.

Agent Interdependencies

Given a joint-action solution trajectory τ and the solution trajectory τ_i for an agent i , we define the following properties about τ and τ_i to formalize the concept of interdependence for the solution trajectory:

For τ , we define *Interdependence* as a pair of actions $(a_i^{t_0+k}, a_j^{t_0})_{i \neq j}$ such that $\text{add}(a_j^{t_0}) \subseteq \text{pre}(a_i^{t_0+k})$. An interdependent pair of actions $(a_i^{t_0+k}, a_j^{t_0})_{i \neq j}$ has two agents, a *Giver* agent performing the action $a_j^{t_0}$ and a *Receiver* agent performing the action $a_i^{t_0+k}$. Each interdependent pair of actions is going to be associated with an object obj_{int} .

We define *Interdependence* as a pair of actions $(a_i^{t_0+k}, a_j^{t_0})_{i \neq j}$ such that $\text{add}(a_j^{t_0}) \subseteq \text{pre}(a_i^{t_0+k})$. For any object in the world and a starting timestep t_0 , the *object influence trajectory* from time t_0 , denoted by $\tau_{\text{obj}}^{t_0}$, captures all state transitions in the plan from timestep t_0 onward where this object is involved.

$$\tau_{\text{obj}}^{t_0} = \{(p_t, a_t, p_{t+1}) \mid t \geq t_0, \exists p \in \text{pre}(a_t) \cup \text{add}(a_t) \cup \text{del}(a_t) \text{ where } \text{obj} \in O(p)\}$$

where $O(p)$ denotes the set of objects mentioned in proposition p . In other words, $\tau_{\text{obj}}^{t_0}$ includes all transitions from timestep t_0 onward where the object explicitly appears in the action's conditions or effects. An interdependence is a *Goal Reaching Interdependence* if the final state of the object associated with that interdependence (obj_{int}) is also present in the set of goal predicates. Let $p_{\text{final}}^{\text{obj}_{\text{int}}}$ denote the final symbolic state of the object associated with the interdependence, corresponding to the last entry in $\tau_{\text{obj}_{\text{int}}}^{t_0}$. Then, the interdependence Int is classified as a Goal Reaching Interdependence if:

$$p_{\text{final}}^{\text{obj}_{\text{int}}} \subseteq p_n^G,$$

where p_n^G is the set of goal propositions at timestep n . Let p_t^{obj} denote the predicate for that object at timestep t , therefore containing information about the state of obj at t . An

interdependence $\text{Int} = (a_i^{t_0+k}, a_j^{t_0})$ associated with object obj_{int} is said to be a *Non-looping Interdependence* if the following conditions hold:

1. The *giver agent* (agent j), who gives the object obj_{int} at timestep t_0 , *does not receive* the object back in the same state at any future timestep $t > t_0 + k$:

$$t > t_0 + k, \quad \text{s.t.} \quad \text{agent } j \text{ receives } \text{obj}_{\text{int}} \text{ in the same state as at time } t_0$$

2. The *receiver agent* (agent i), who receives the object at timestep $t_0 + k$, *did not have* the object in that same state at any time $t < t_0 + k$:

$$t < t_0 + k, \quad \text{s.t.} \quad \text{agent } i \text{ had } \text{obj}_{\text{int}} \text{ in the same state}$$

An interdependence $\text{Int} = (a_i^k, a_j^{k-t})$ is a **Constructive Interdependence**, if it is a *Goal Reaching Interdependence* and a *Non-looping Interdependence*. Consider a scenario in the Counter Circuit layout where agent j places an onion on the counter at timestep t_0 via action $a_j^{t_0}$, whose effect is $\text{add}(a_j^{t_0}) = \{\text{onion-on-counter}\}$. Subsequently, at timestep $t_0 + k$, agent i performs action $a_i^{t_0+k}$ to pick up the onion from the counter, with precondition $\text{pre}(a_i^{t_0+k}) = \{\text{onion-on-counter}\}$. This pair of actions $(a_i^{t_0+k}, a_j^{t_0})$ constitutes a *sequential interdependence* Int linked to the object $\text{obj}_{\text{int}} = \text{onion}$. The associated object influence trajectory $\tau_{\text{onion}}^{t_0}$ captures all state transitions involving the onion, culminating in a final state where the soup contains the onion. Provided that the onion is not returned to agent j in the same state and that agent i had not previously held the onion in that state, this interdependence is *Non-looping*. Consequently, this interaction qualifies as a Constructive Interdependence. A *Trigger* action for an agent is placing the onion on the counter, since it could potentially be the precondition for the other agent picking that onion from the counter.

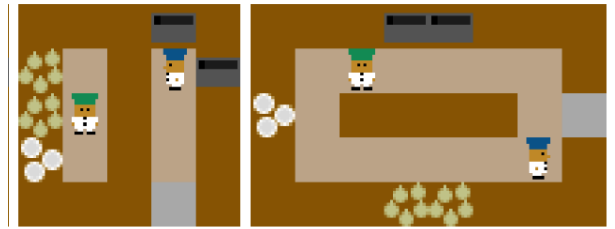


Figure 2: Left: Forced coordination layout which is a required cooperation (RC) setting. Right: Counter circuit layout which is a non-required cooperation (Non-RC) setting.

Experiments

We evaluate state-of-the-art (SOTA) zero-shot cooperation (ZSC) agents in human-AI teams using the proposed metric within a benchmark environment. Our analysis focuses on these questions:

- *RQ1*: Can ZSC agents recognize and adapt to partners who initiate cooperative strategies?
- *RQ2*: To what extent do teams with ZSC agents and human partners exhibit cooperative behavior in Non-Required Cooperation (Non-RC) scenarios?
- *RQ3*: How does cooperative behavior differ between RC and Non-RC settings when ZSC agents team with humans?

Metrics: We quantify cooperation in teams using two key metrics: constructive interdependencies (Int_{cons}) which capture instances where agents’ actions meaningfully contribute to task completion through interdependent actions and non-constructive interdependencies ($Int_{\text{non-cons}}$) which quantifies unproductive or redundant interdependencies between agents’ actions. We also measure how many interdependencies are initiated by the teammates ($\%P_{\text{tot-sub}}^{\text{trig}}$), and of those, how many are *not accepted and acted upon* by the ZSC agents ($\%P_{\text{trig}}^{\text{not-trig-acc}}$).

Environment: The environment is a fully observable, timed gridworld where a team of two players must collaboratively prepare and deliver three soups within a fixed time limit. It contains onion dispensers, dish dispensers, pots, serving stations, and empty counters. Players act concurrently at each timestep and can either move or interact with objects. To complete a soup and get a reward, the team must collect and place three onions in a pot, wait for it to cook, retrieve a dish, transfer the soup, and deliver it. Each player and counter can hold only one object at a time. Task rewards are shared between both players upon successful delivery, which is *expected to incentivize efficient collaboration*.

SOTA Methods: FCP (Strouse et al. 2022), MEP (Zhao et al. 2022), HSP (Yu et al. 2023), and COLE (Li et al. 2024) utilize a two-stage training pipeline. First, a diverse population of partners is generated through self-play. Next, an ego agent is trained via reinforcement learning by interacting with sampled partners from this population, optimizing primarily for episodic task reward. This task reward, which serves as the training signal, is also the main evaluation metric used to assess cooperation when these agents are paired with previously unseen teammates, including humans.

Evaluation Partners: Using the proposed metric, we analyze the cooperative behavior of ZSC agents, on the forced coordination (RC) and counter circuit (non-RC) layout (Fig. 2), in the following settings:

1. *Scripted Cooperative Agent:* Agents are paired with a scripted partner executing a fixed cooperative strategy. This setting tests whether ZSC agents can recognize and adapt to consistently cooperative behaviors (*RQ1*).
2. *Self-Play:* Agents are paired with an identical copy of themselves. Serving as a baseline, this evaluates their ability to coordinate when interacting with the same policy (*RQ2*).
3. H_{proxy} : Agents interact with learned human behavior models (Wang et al. 2024a). Since scaling user studies is costly, this provides an additional proxy for assessing agent cooperation with humans. (*RQ2*, *RQ3*)
4. *Human Teammates:* We conduct a user study pairing ZSC agents with 36 human participants to measure

cooperative behavior in real-world human-agent teams (*RQ1, RQ2, RQ3*).

User Study: We recruited 36 participants from our university in the range from 18 to 31 who were pursuing either an undergraduate or a graduate degree. We conducted a pilot study on 5 participants spread across each of the two evaluation domains. The final study had a sample size of 31 participants. Participants had an average age of 20.75 years, and a median age of 22.5 years. Out of the 36, there were 24 male participants and 12 female participants. 23 participants (63.9%) reported to not have any familiarity with playing the Overcooked game earlier, and the remaining 13 (36.1%) were familiar with the game.

Results

In this section, we present the evaluation of ZSC agents paired with various partners. Using our proposed metric, we quantify the cooperative behavior exhibited by these teams across different settings.

Agent	Task Reward	Int_{cons}	$Int_{\text{non-cons}}$	$Int_{\text{cons}}^{\text{coop}}$
COLE	36	0.60	1.20	4.425
MEP	43.33	0.83	1.67	6.59
HSP	0	0	0.50	0
FCP	0	0	0.17	0

(a) Average rewards and cooperation metrics across 50 runs. $Int_{\text{cons}}^{\text{coop}}$ represents the interdependence if both agents followed the coordination policy.

Agent	$\%P_{\text{tot-sub}}^{\text{trig}}$	$\%P_{\text{trig}}^{\text{not-trig-acc}}$
COLE	58.5	88.88
MEP	41.28	75.55
HSP	58.57	100.0
FCP	41.79	100.0

(b) Average response rates of ZSC agents to coordination attempts. $\%P_{\text{tot-sub}}^{\text{trig}}$ shows the percentage of interdependencies triggered by the scripted agent, while $\%P_{\text{trig}}^{\text{not-trig-acc}}$ indicates the percentage of triggered interdependencies that were not accepted by the ZSC agents.

Table 1: Analysis of ZSC agents’ performance when paired with a cooperative scripted partner in Non-RC layout. Blue cells highlight cases with high task reward despite much lower cooperation than the joint coordination policy; red cells mark frequently triggered interdependencies by human which are frequently rejected by ZSC agents.

ZSC Agent paired with Cooperative Partner

We assess whether ZSC agents can recognize and adapt to a partner employing a known coordination policy by quantifying the resulting coordination using our proposed metric.

Specifically, we evaluate teams in the Counter Circuit layout, implementing the coordination strategy introduced by Carroll et al. (2020), as illustrated in Fig. 1. In this strategy, the green-hat chef places onions on the counter, while the blue-hat chef picks them up, places them in the pot, and later serves the soup. Our experimental setup assigns the scripted agent to the green-hat role and the ZSC agent to the blue-hat role, testing the agent’s ability to complement and adapt to its partner’s cooperative behavior.

The joint coordination policy exhibits sequential interdependence, which we capture using the constructive interdependency metric (Int_{cons}). We also measure constructive interdependencies when both agents strictly follow the joint coordination policy (Int_{cons}^{coop}). This represents the upper bound on coordination achievable if the ZSC agent fully adapts to the scripted partner. Comparing these values enables us to evaluate the extent to which ZSC agents successfully recognize and complement cooperative strategies in a Non-RC setting.

From Table 1a, we observe that all ZSC agents exhibit low levels of constructive interdependence when paired with a scripted partner following a coordination policy. For example, COLE achieves a task reward of 36 but only reaches $Int_{cons} = 0.60$, far below the cooperative upper bound of $Int_{cons}^{coop} = 4.425$. Similarly, MEP obtains the highest reward of 43.33 among the agents but still falls short in coordination, with $Int_{cons} = 1.83$ compared to $Int_{cons}^{coop} = 6.59$. As shown in Table 1b, the scripted agent frequently attempts to initiate interdependencies ($\%P_{tot-sub}^{trig}$ of 58.5 for COLE and 41.28 for MEP), yet these are often ignored by the ZSC agents, with rejection rates of 88.88% and 75.55% respectively. Despite the partner’s consistent cooperative behavior, HSP and FCP entirely fail to coordinate, with zero constructive interdependence and 100% rejection rates.

Task vs Teaming Performance of ZSC in Self-Play

We evaluate whether agents can induce cooperative strategies when paired with identical copies of themselves (Table 2). We find that while task rewards are generally high, the level of constructive interdependence varies widely. For example, COLE and HSP both achieve a reward of 120 in non-RC, but differ substantially in Int_{cons} (10.23 vs. 1.82), indicating that similar task outcomes can arise from vastly different levels of coordination. In RC layouts, however, we observe a strong alignment between task reward and Int_{cons} (e.g., COLE: 200 and 30.43; MEP: 140 and 29.5). Furthermore, across all non-RC cases, $Int_{non-cons}$ consistently exceeds Int_{cons} , as seen in MEP (10.19 vs. 1.05) and HSP (12.32 vs. 1.82), indicating that when interdependencies do arise, they are often unproductive or misaligned.

Task vs Teaming Performance of ZSC Agents Paired with H_{proxy}

As observed in Table 3, constructive interdependencies remain low in Non-RC layout, indicating weak cooperation when ZSC agents are paired with H_{proxy} . For example, MEP achieves the highest task reward in Non-RC (140) but exhibits the lowest Int_{cons} (1.05), highlighting a decou-

Agent	Task Reward		Int_{cons}		$Int_{non-cons}$	
	Non-RC	RC	Non-RC	RC	Non-RC	RC
COLE	120	200	10.23	30.43	6.58	6.83
MEP	100.00	140	1.05	29.5	10.19	7.53
HSP	120	100	1.82	22.87	12.32	10.667
FCP	60	80	0	16.0	0	16.0

Table 2: Average rewards and cooperation metrics of ZSC agents in self-play across Non-RC and RC layouts. Green cells highlight correlation of reward and cooperation in RC; blue cells indicate high task reward despite low cooperation in Non-RC; red cells mark non-constructive interdependencies.

Agent	Task Reward		Int_{cons}		$Int_{non-cons}$	
	Non-RC	RC	Non-RC	RC	Non-RC	RC
COLE	100	200	7.52	35.00	10.98	3.32
MEP	140.00	180	1.05	29.5	12.21	9.53
HSP	160	180	4.5	30.00	11.32	7.667
FCP	100	120	6.021	24	15.02	4.0

Table 3: Average rewards and cooperation metrics for ZSC agents when paired with H_{proxy} . Green cells highlight correlation of reward and cooperation in RC; blue cells indicate high task reward despite low cooperation in Non-RC; red cells mark non-constructive interdependencies.

pling between reward and cooperative behavior. By contrast, in RC layout, we observe a marked improvement in constructive interdependencies. COLE, for instance, attains both the highest task reward (200) and the highest Int_{cons} (35), demonstrating strong alignment between cooperation and performance. This pattern is consistent across agents—highlighting the inadequacy of task reward as a proxy for teamwork in Non-RC settings.

Task vs Teaming Performance of ZSC Agents paired with Human Teammates

From Table 4a, constructive interdependencies (Int_{cons}) remain consistently low in Non-RC layouts even with human partners. For example, COLE and HSP achieve reasonably high task rewards (76.21 and 41.11) but low Int_{cons} (1.89 and 1.39 respectively). This trend persists among the best-performing teams (Table 4b), where MEP achieves a reward of 80 but with only 1.285 constructive interdependencies, underscoring the limited cooperation despite successful task completion. In the RC layout, we observe a strong positive correlation between task performance and constructive interdependencies. For instance, the HSP agent achieves the highest task reward of 120 when paired with a human participant, and this is accompanied by the highest number of constructive interdependencies (25.00) as shown in Table 4b. Thus, task reward can serve as a reliable proxy

Agent	Task Reward		Int _{cons}		Int _{non-cons}	
	Non-RC	RC	Non-RC	RC	Non-RC	RC
COLE	76.21	56.87	1.89	11.37	6.29	2.87
MEP	50	44.1	0.92	8.69	1.28	2.76
HSP	41.11	60.55	1.38	12.05	2.13	3.08
FCP	22.55	35.34	0.97	7.06	0.872	3.44

(a) Average rewards and cooperation metrics for ZSC agents across all runs with human participants.

Agent	Task Reward		Int _{cons}		Int _{non-cons}	
	Non-RC	RC	Non-RC	RC	Non-RC	RC
COLE	120	100	3.79	20.00	3.0	0.33
MEP	80.00	120	1.285	24.0	7.47	0
HSP	80	120	3.5	25.00	1.25	1.667
FCP	60	100	3.667	20	1.334	3.0

(b) Rewards and cooperation metrics for top-performing human-agent teams for each ZSC agent.

Agent	%P _{tot-sub} ^{trig}		%P _{trig} ^{not trig-acc}	
	Non-RC	RC	Non-RC	RC
COLE	60.28	45.28	70.05	38.34
MEP	66.82	43.57	82.39	39.82
HSP	52.22	42.92	80.85	40.58
FCP	58.30	43.62	98.41	36.84

(c) Response rates of ZSC agents to coordination attempts by human teammates. %P_{tot-sub}^{trig} shows the percentage of interdependencies triggered by the humans, while %P_{trig}^{not trig-acc} indicates the percentage of triggered interdependencies that were not accepted by the ZSC agents.

Table 4: Comprehensive analysis of human-agent team performance across Non-RC and RC settings. Green cells highlight correlation of reward and cooperation in RC; blue cells indicate high task reward despite low cooperation in Non-RC; red cells mark frequently triggered interdependencies by human which are frequently rejected by ZSC agents; yellow cells mark non-constructive interdependencies.

for cooperative behavior in RC settings, in contrast to non-RC layouts where such alignment breaks down. Table 4c shows that humans frequently initiate cooperative interactions (e.g., MEP triggers 66.82% interdependencies in Non-RC), but ZSC agents often reject these attempts. We also found a strong positive correlation in RC settings (Pearson $r = 0.81$, $p < 0.001$), and weak correlation in Non-RC settings, (Pearson $r = 0.19$), confirming that agents can achieve high reward without meaningful cooperation. Using the proposed metric, our analysis of human-agent

teams reveals the following key insights:

- When paired with partners—scripted or human—who attempt to initiate cooperation, ZSC agents fail to adapt to their partners and frequently reject coordination attempts.
- In Non-RC layouts, ZSC agents can achieve high task performance, but they do not induce cooperative behavior. Agents that perform well demonstrate significantly low levels of constructive interdependence.
- In RC layouts, task reward aligns much more strongly with inter-agent coordination. Agents that perform well also demonstrate higher levels of constructive interdependence.
- In cases where interdependence emerges in Non-RC settings, these are frequently non-constructive, indicating misaligned or ineffective coordination.

These results reveal the inadequacy of task reward as a standalone metric for evaluating cooperation in human-agent teams in Non-RC settings. This work highlights a critical gap in current state-of-the-art for Zero-Shot Coordination: their limited ability to engage in meaningful cooperation when paired with partners attempting to coordinate.

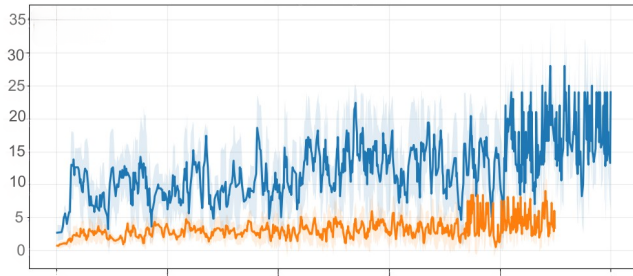
Incorporating Teaming Reward into Training

Motivated by this gap, we conducted preliminary experiments incorporating the proposed cooperation metric into the training of SOTA algorithms. Concretely, we modify the training reward as follows, where r_{task} is the task reward and $r_{teaming}$ is the teaming reward measured by constructive interdependencies: $\mathbf{r}_{modified} = r_{task} + \alpha \times r_{teaming}$, $\mathbf{r}_{original} = r_{task}$.

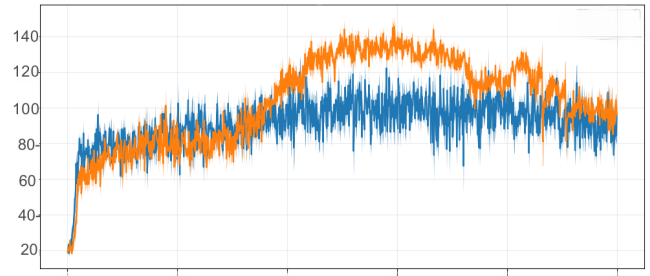
We train SOTA ZSC agents under both reward formulations and compare their learning dynamics. From Fig. 3, we observe that agents trained with the modified reward consistently increase their teaming reward throughout training, indicating that they learn to engage in behaviors that lead to constructive interdependencies. In contrast, agents trained with the original reward show nearly stagnant teaming reward, confirming that standard training does not incentivize cooperative interactions. Importantly, we observe that task performance remains comparable between the modified and original training rewards, suggesting that incorporating interdependence does not significantly compromise task reward. Overall, these findings show that incorporating the proposed cooperation metric directly into the learning objective enables SOTA ZSC agents to learn cooperative interactions without compromising on the task reward, addressing a key limitation identified in our evaluation.

Conclusion

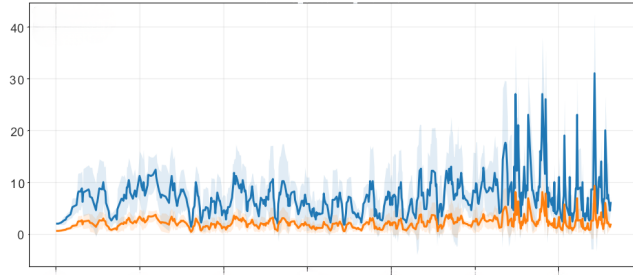
In this work, we propose a metric to evaluate cooperation in human-AI teams, addressing key limitations of task-reward-based evaluations in zero-shot cooperation (ZSC). Our findings reveal that current state-of-the-art ZSC agents often achieve high task rewards without engaging in meaningful coordination with their partners. Through controlled evaluation with a scripted agent, human proxy, and real human participants, we showed that these agents fail to recognize or



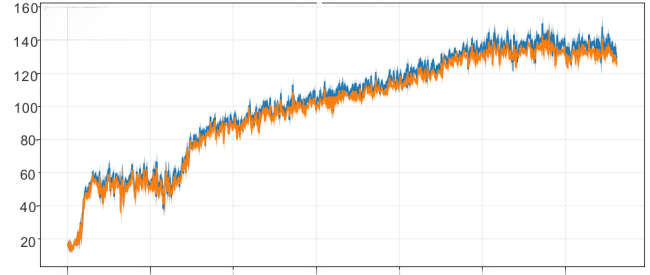
(a) FCP: Teaming Reward



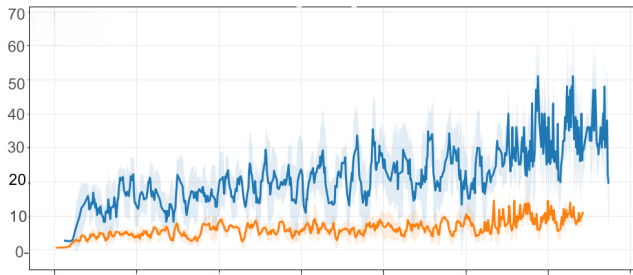
(b) FCP: Task Reward



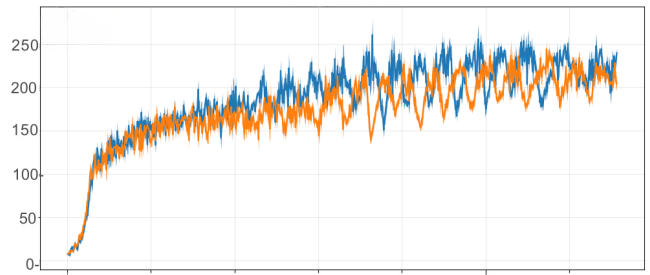
(c) MEP: Teaming Reward



(d) MEP: Task Reward



(e) HSP: Teaming Reward



(f) HSP: Task Reward

Figure 3: Training curves comparing agents trained with the original task-only reward (orange curves) and the modified reward (blue curves) that incorporates interdependence ($\alpha = 0.3$). All agents are trained for 5 million timesteps. The orange curves correspond to training with $\mathbf{r}_{original}$, while the blue curves correspond to training with $\mathbf{r}_{modified}$. Across all three algorithms (FCP, MEP, HSP), adding the teaming reward leads to a consistent increase in constructive interdependencies, while task performance remains unchanged.

respond appropriately to cooperative signals, particularly in environments where cooperation is optional. This highlights that task reward alone is an insufficient proxy for cooperative behavior in scenarios where cooperation is optional, emphasizing the need of explicit metrics to accurately quantify inter-agent coordination. Future work would include broadening this metric to include other kinds of interdependencies and cooperative behaviors. Another research direction is to use the interdependence metric as an additional reward signal to guide learning towards effective cooperation, rather than relying on coordination to emerge implicitly from the task reward. In non-RC settings, the *shadowed equilibrium*

problem (Matignon, Laurent, and Le Fort-Piat 2012; Fulda and Ventura 2007) causes agents to not explore the cooperative strategies during training, since multiple equilibria exist including the non-cooperative strategies. We also provide preliminary experiments showing that integrating the interdependence metric as a reward signal encourages agents to actively recognize and pursue coordination during exploration, potentially learning to play with a diverse set of partners and reducing miscoordination in human-agent teaming scenarios. Ultimately, this work paves the way for developing ZSC agents that not only succeed at task performance but also robustly cooperate with diverse human behaviors.

Acknowledgments

This research is supported primarily by ONR grant N0001423-1-2409. This research is also supported by DARPA grant HR00112520016, and gifts from Qualcomm and Amazon. Special thanks to Dhanush Giriyan for guidance and technical support

References

- Barton, S. L.; Waytowich, N. R.; and Asher, D. E. 2018. Coordination-driven Learning in Multi-agent Problem Spaces. In *AAAI Fall Symposium: ALEC*.
- Barton, S. L.; Waytowich, N. R.; Zaroukian, E.; and Asher, D. E. 2018. Measuring collaborative emergent behavior in multi-agent reinforcement learning. arXiv:1807.08663.
- Bishop, J.; Burgess, J.; Ramos, C.; Driggs, J.; Williams, T.; Tossell, C.; Phillips, E.; Shaw, T.; and de Visser, E. 2020. CHAOPT: A Testbed for Evaluating Human-Autonomy Team Collaboration Using the Video Game Overcooked!2. 1–6.
- Cai, C. J.; Reif, E.; Hegde, N.; Hipp, J.; Kim, B.; Smilkov, D.; Wattenberg, M.; Viegas, F.; Corrado, G. S.; Stumpe, M. C.; and Terry, M. 2019. Human-Centered Tools for Coping with Imperfect Algorithms during Medical Decision-Making. arXiv:1902.02960.
- Carroll, M.; Shah, R.; Ho, M. K.; Griffiths, T. L.; Seshia, S. A.; Abbeel, P.; and Dragan, A. 2020. On the Utility of Learning about Humans for Human-AI Coordination. arXiv:1910.05789.
- Fikes, R. E.; and Nilsson, N. J. 1971. STRIPS: a new approach to the application of theorem proving to problem solving. In *Proceedings of the 2nd International Joint Conference on Artificial Intelligence, IJCAI'71*, 608–620. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Fontaine, M. C.; Hsu, Y.-C.; Zhang, Y.; Tjanaka, B.; and Nikolaidis, S. 2021. On the Importance of Environments in Human-Robot Coordination. arXiv:2106.10853.
- Fulda, N.; and Ventura, D. 2007. Predicting and preventing coordination problems in cooperative Q-learning systems. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI'07*, 780–785. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Ghavamzadeh, M.; Mahadevan, S.; and Makar, R. 2006. Hierarchical multi-agent reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 13(2): 197–229.
- Johnson, M.; Bradshaw, J. M.; Feltovich, P. J.; Jonker, C. M.; van Riemsdijk, M. B.; and Sierhuis, M. 2014. Coactive design: designing support for interdependence in joint activity. *J. Hum.-Robot Interact.*, 3(1): 43–69.
- Johnson, M.; Vignatti, M.; and Duran, D. 2020. *Understanding Human-Machine Teaming through Interdependence Analysis*, 209–233. ISBN 9780429459733.
- Knott, P.; Carroll, M.; Devlin, S.; Ciosek, K.; Hofmann, K.; Dragan, A. D.; and Shah, R. 2021. Evaluating the Robustness of Collaborative Agents. arXiv:2101.05507.
- Lerer, A.; and Peysakhovich, A. 2019. Learning Existing Social Conventions via Observationally Augmented Self-Play. arXiv:1806.10071.
- Li, Y.; Zhang, S.; Sun, J.; Zhang, W.; Du, Y.; Wen, Y.; Wang, X.; and Pan, W. 2024. Tackling Cooperative Incompatibility for Zero-Shot Human-AI Coordination. arXiv:2306.03034.
- Liang, C.; Proft, J.; Andersen, E.; and Knepper, R. A. 2019. Implicit Communication of Actionable Information in Human-AI teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19*, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.
- Lopez, J.; Textor, C.; Lancaster, C.; Schelble, B.; Freeman, G.; Zhang, R.; McNeese, N.; and Pak, R. 2023. The complex relationship of AI ethics and trust in human-AI teaming: insights from advanced real-world subject matter experts. *AI and Ethics*, 4: 1213–1233.
- Lou, X.; Guo, J.; Zhang, J.; Wang, J.; Huang, K.; and Du, Y. 2023. PECAN: Leveraging Policy Ensemble for Context-Aware Zero-Shot Human-AI Coordination. arXiv:2301.06387.
- Ma, L. M.; Ijtsma, M.; Feigh, K. M.; and Pritchett, A. R. 2022. Metrics for Human-Robot Team Design: A Teamwork Perspective on Evaluation of Human-Robot Teams. *J. Hum.-Robot Interact.*, 11(3).
- Matignon, L.; Laurent, G. J.; and Le Fort-Piat, N. 2012. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. *The Knowledge Engineering Review*, 27(1): 1–31.
- Moran, S.; Pantidi, N.; Bachour, K.; Fischer, J. E.; Flintham, M.; Rodden, T.; Evans, S.; and Johnson, S. 2013. Team reactions to voiced agent instructions in a pervasive game. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces, IUI '13*, 371–382. New York, NY, USA: Association for Computing Machinery. ISBN 9781450319652.
- Nalepka, P.; Gregory-Dunsmore, J.; Simpson, J.; Patil, G.; and Richardson, M. 2021. Interaction Flexibility in Artificial Agents Teaming with Humans.
- Pateria, S.; Subagdja, B.; and Tan, A.-H. 2019. Multi-agent Reinforcement Learning in Spatial Domain Tasks using Inter Subtask Empowerment Rewards. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 86–93.
- Ries, A. J.; Aroca-Ouellette, S.; Roncone, A.; and de Visser, E. 2024. Gaze-informed Signatures of Trust and Collaboration in Human-Autonomy Teams. *ArXiv*, abs/2409.19139.
- Sarkar, B.; Shih, A.; and Sadigh, D. 2023. Diverse Conventions for Human-AI Collaboration. arXiv:2310.15414.
- Singh, R.; Miller, T.; and Sonenberg, L. 2014. A Preliminary Analysis of Interdependence in Multiagent Systems. volume 8861. ISBN 978-3-319-13190-0.
- Singh, R.; Miller, T.; and Sonenberg, L. 2016. Communication and Shared Mental Models for Teams Performing Interdependent Tasks. ISBN 9783319665948.
- Strouse, D.; McKee, K. R.; Botvinick, M.; Hughes, E.; and Everett, R. 2022. Collaborating with Humans without Human Data. arXiv:2110.08176.

Textor, C.; Zhang, R.; Lopez, J.; Schelble, B. G.; McNeese, N. J.; Freeman, G.; Pak, R.; Tossell, C.; and de Visser, E. J. 2022. Exploring the Relationship Between Ethics and Trust in Human–Artificial Intelligence Teaming: A Mixed Methods Approach. *Journal of Cognitive Engineering and Decision Making*, 16(4): 252–281.

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

Wang, M.; Wu, J.; Ma, S.; Li, N.; Zhang, P.; Gu, N.; and Lu, T. 2025. Adaptive Human-Agent Teaming: A Review of Empirical Studies from the Process Dynamics Perspective.

Wang, X.; Zhang, S.; Zhang, W.; Dong, W.; Chen, J.; Wen, Y.; and Zhang, W. 2024a. ZSC-Eval: An Evaluation Toolkit and Benchmark for Multi-agent Zero-shot Coordination. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Wang, Z. M.; Peng, Z.; Que, H.; Liu, J.; Zhou, W.; Wu, Y.; Guo, H.; Gan, R.; Ni, Z.; Yang, J.; Zhang, M.; Zhang, Z.; Ouyang, W.; Xu, K.; Huang, S. W.; Fu, J.; and Peng, J. 2024b. RoleLLM: Benchmarking, Eliciting, and Enhancing Role-Playing Abilities of Large Language Models. arXiv:2310.00746.

Wu, H.; Ghadami, A.; Bayrak, A. E.; Smereka, J. M.; and Epureanu, B. I. 2023. Evaluating Emergent Coordination in Multi-Agent Task Allocation Through Causal Inference and Sub-Team Identification. *IEEE Robotics and Automation Letters*, 8(2): 728–735.

Yu, C.; Gao, J.; Liu, W.; Xu, B.; Tang, H.; Yang, J.; Wang, Y.; and Wu, Y. 2023. Learning Zero-Shot Cooperation with Humans, Assuming Humans Are Biased. arXiv:2302.01605.

Zhang, S.; Wang, X.; Zhang, W.; Chen, Y.; Gao, L.; Wang, D.; Zhang, W.; Wang, X.; and Wen, Y. 2024. Mutual Theory of Mind in Human-AI Collaboration: An Empirical Study with LLM-driven AI Agents in a Real-time Shared Workspace Task. arXiv:2409.08811.

Zhang, Y.; Sreedharan, S.; and Kambhampati, S. 2016. A formal analysis of required cooperation in multi-agent planning. In *Proceedings of the Twenty-Sixth International Conference on International Conference on Automated Planning and Scheduling*, ICAPS’16, 335–343. AAAI Press. ISBN 1577357574.

Zhao, M.; Simmons, R.; and Admoni, H. 2022. Coordination with Humans via Strategy Matching. arXiv:2210.15099.

Zhao, R.; Song, J.; Yuan, Y.; Haifeng, H.; Gao, Y.; Wu, Y.; Sun, Z.; and Wei, Y. 2022. Maximum Entropy Population-Based Training for Zero-Shot Human-AI Coordination. arXiv:2112.11701.