Evaluating Emergent Coordination in Multi-Agent Task Allocation Through Causal Inference and Sub-Team Identification

Haochen Wu, Amin Ghadami, Alparslan Emrah Bayrak, Jonathon M. Smereka, and Bogdan I. Epureanu

***Abstract*—Coordination in multi-agent systems is a vital component in teaming effectiveness. In dynamically changing situations, agent decisions depict emergent coordination strategies from following pre-defined rules to exploiting incentive-driven policies. While multi-agent reinforcement learning shapes team behaviors from experience, interpreting learned coordination strategies offers benefits in understanding complex agent dynamics and further improvement in developing adaptive strategies for evolving and unexpected situations. In this work, we develop an approach to quantitatively measure team coordination by collecting decision time series data, detecting causality between agents, and identifying statistically high coordinated sub-teams in decentralized multi-agent task allocation operations. We focus on multi-agent systems with homogeneous agents and homogeneous tasks as the strategy formation is more ambiguous and challenging than heterogeneous teams with specialized capabilities. Emergent team coordination is then analyzed using rule-based and reinforcement learning-based strategies for task allocation in operations at different demand stages (stress) levels. We also investigate correlation vs. causation and agent over- or under-estimating demand levels.**

***Index Terms*—Multi-Robot Systems, Reinforcement Learning, Planning, Scheduling and Coordination, Task Planning.**

I. Introduction

IN REAL-WORLD scenarios involving multi-agent teams, coordinated strategies are naturally formed and driven by a common goal and individual preferences, without explicit communication of intended actions. Advanced technologies in collaborative autonomy have significantly enhanced the effectiveness of multi-agent teams including swarm robots and human-machine teams. Understanding complex interactions, behavior correlations, causality, and evolving strategies of multi-agent teams remains challenging. Coordination in multi-agent task allocation often considers centralized consensus among distributed systems through game theory and optimization [1] or utilizes deep reinforcement learning techniques [2] to train polices under uncertainties by interacting with the environment. To investigate multi-agent coordination strategies, inverse reinforcement learning techniques [3], [4], [5] enable learning from human demonstrations, and explainable reinforcement learning aims to create interpretable reward functions for agent behaviors. There is a need, however, to quantify the synchronization and synergy of teaming behaviors under different operations to gain insights into the underlying team dynamics and the emergent coordination strategies.

Multi-agent reinforcement learning (MARL) allows agents to learn coordinated behaviors, where team strategies evolve from random exploratory to objective-oriented exploitative behaviors [6], [7]. In cooperative and competitive tasks, rewards are used to find the correlations between agent behaviors [8], and population-based methods [9] select team members by competition. Sometimes high-performing MARL trained teams in predator-prey systems do not reach the desired level of collaboration [10], especially in large teams [11]. Although correlation in rewards and performance metrics help illustrate the evolving behaviors during learning, the coordination patterns depicted by agent decisions have not been studied, and whether the suitable strategy is implemented facing unexpected situations is therefore not guaranteed. Studying coordination allows analyzing and determining team strategies in various situations by directly using operation demands and agent behaviors.

To analyze decisions made by multi-agent teams and the cause-and-effect relationships among these decisions, it is important to visualize and measure whether agent decision patterns can be informed by past decisions of other agents, and whether certain agents form sub-teams that are highly synchronized compared to others. In this paper, we utilize the concepts of decision causality [12] and bidirectional weighted networks [13] for evaluating coordination in decentralized multi-agent teams operating in large-scale task allocation problems. Measuring the level of team coordination and identifying highly coordinated sub-teams are crucial for understanding emergent strategies and recognizing the need for them in evolving situations and

Manuscript received 8 August 2022; accepted 16 December 2022. Date of publication 22 December 2022; date of current version 29 December 2022. This letter was recommended for publication by Associate Editor M. Meghjani, and Editor T. Asfour upon evaluation of the reviewers' comments. This work was supported in part by Automotive Research Center, University of Michigan and in part by DISTRIBUTION A: Approved for Public Release; Distribution Unlimited; OPSEC# 6703. *(Corresponding author: Haochen Wu.)*

Haochen Wu, Amin Ghadami, and Bogdan I. Epureanu are with the University of Michigan, Ann Arbor, MI 48109 USA (e-mail: haochenw@umich.edu; aghadami@umich.edu; epureanu@umich.edu).

Alparslan Emrah Bayrak is with the School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: ebayrak@stevens.edu).

Jonathon M. Smereka is with the US Army CCDC Ground Vehicle Systems Center, Warren, MI 48397 USA (e-mail: jonathon.m.smereka.civ@mail.mil).

Digital Object Identifier 10.1109/LRA.2022.3231497

unforeseen events, especially when agents are facing extreme workload and stress, and when agents over- or under-estimate the task demand. We demonstrate the developed metrics for quantifying team coordination in multi-agent systems operating in an environment with predefined centralized rule-based strategies and decentralized team strategies trained by MARL. We also show that RL-trained decentralized multi-agent teams can learn not only effective operation completion but also coordinated behaviors, compared to rule-based policies.

The key contributions of this work can be summarized as:

- Developing a generalized approach to measure causality among agents and identify coordinated sub-teams using time series data of agents' decisions in decentralized multi-agent teams,
- Defining a quantitative metric for emergent coordination to acquire patterns in rule-based and reward-driven strategies across various environment demand levels,
- Investigating the effect of over- and under-estimating task demands by agents on team coordination and performance.

To our knowledge, this work is the first attempt to apply causal inference for evaluating coordination behaviors in multi-agent task allocation, which is capable of quantitatively investigating agent-level coordination, sub-team formation, and team-level strategies.

II. Related Work

Training agents to learn policies for coordination and competition has shown promising effectiveness using MARL [7], [8], [9]. Studies have involved the construction of coordination graphs [14] and distributed networks [15] using deep learning techniques [16] to learn the quality values for the joint action of each agent pair, and demonstrated the benefit of using correlation in reward histories on learning team strategies [8]. Such learning methods utilize the underlying coordination structure to guide the learning process, but they have not focused on analyzing the learned strategy or coordination level across different teams or situations.

To address the issue of black box decisions made by artificial intelligence, explainable reinforcement learning [17], [18], [19] tends to evaluate the interestingness elements in agent decisions such as frequency, execution uncertainty, and favorable/adverse situations [20]. Several studies have investigated learned formation strategies [21] and path planning strategies using heat maps [22]. However, theses approaches have not explained the correlation or causality effect between agents, and they are not general enough to merely use state and decision information. Moreover, it would become challenging to study task allocation problems when multiple strategies lead to the optimal team performance.

The Pearson correlation [23] and KL divergence [24] are widely used to measure the relationship between two random variables, to match the probability of the desired strategy [25], or to identify strategy switching [26], but these metrics produce a symmetric correlation between variables and do not address the fact that correlation does not imply causation. In economics and ecosystems, statistical tests including Granger Causality [27] and Convergence Cross Mapping (CCM) [12] have been developed to detect the causality between time series of two variables. While Granger Causality assumes that variables are separable, CCM variants [28], [29], [30], [31] seek synergistic effects between variables in stochastic and dynamical systems. Causality has been rarely studied in multi-agent task allocation to quantitatively measure the level of coordination in strategies.

The novelty of our approach includes the application of asymmetric causation technique (CCM) in task allocation and the extension from two-agent causation to quantification of multi-agent coordination by identifying synergistic sub-teams using community finding techniques [32].

III. Methods

In the context of multi-task multi-agent decision-making problems, the team coordination operation is considered as a complex simulated scenario where tasks, represented by demand levels, are to be accomplished in multiple time steps by agents equipped with various task-handling capability levels. The operation is accomplished once the demand levels of all tasks are reduced to zero. With the presence of uncertainties, the dynamics of the task demand levels are stochastic and dependent on each other and/or the agents' capability levels at the assigned tasks. This flexible formulation of the environment is applicable in various types of scenarios ranging from homogeneous-task homogeneous-agent with one-step task assignment (i.e., predator swarm robots) to heterogeneous-task heterogeneous-agent with multi-step dynamic task allocation (i.e., human-autonomy teaming). In such an environment, agents are incentivized to accomplish tasks within fewer time steps. The coordination strategy is emerging from cooperating on the same task to simultaneously working on different tasks for efficient teaming. Quantitative analysis of team coordination is necessary for understanding emergent teaming behaviors and varying strategies under changing situations. Therefore, we propose a method that utilizes bidirectional causality in decisions and coordinated sub-team identification to evaluate team coordination.

A. Preliminaries

1) Formulation: Dec-POMDP: The environment is formulated as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP) [33], which is defined as a tuple $< S, \{A_i\}, \mathcal{T}, R, \{Z_i\}, \mathcal{O}, \gamma >$, where i is the agent index and $s \in S$ is the actual state of the environment. In the context of task allocation, states are represented as task demand levels. At each time step, each agent i receives partial observation $z_i \in Z_i$ (i.e., the demand level of the assigned task), and the joint observation is $z = \{z_i\}$. Then, each agent i makes a task allocation decision and executes action $a_i \in A_i$, and the joint action is $a = \{a_i\}$. With actions taken, the current environment state s transitions to the next state s' with probability $\mathcal{T} = Pr(s'|s, a)$. Agents receive reward $r = f_R(s, a)$ and joint observation $z = \{z_i\}$ with probability $\mathcal{O} = Pr(z|s,' a)$ for making decisions in the next step. In an episode of h time steps, agents follow policy π to make decisions. The cumulative discounted reward

is $R_c = \sum_{t=1}^{h} \gamma^{t-1} f_R(s^t, a^t)$, where $\gamma \in [0, 1)$ is a discount factor introduced to provide more reward at earlier time steps.

2) Decision Strategy: Reinforcement Learning vs. Rules: By interacting with the environment, agents associate their past experience with the reward they received. RL-trained methods allow agents to learn the optimal policy that maximizes the cumulative discounted reward R_c, but the learned behaviors are often hard to explain [20]. In contrast, rule-based decisions are comparably interpretable. For example, the rule could be simply forming groups of three to cooperate for addressing highly demanding tasks and individually executing different low-demand tasks. However, these strategies may not always be optimal or effective.

3) Convergence Cross Mapping as Coordination: When two agents express coordinated behaviors, the behavior of one agent is predictable from the other's. Thus, two agents are coordinating not only when they work on the same task, but also when the next decision of one agent is predictable by observing the past decisions of the other agent. Such predictability is considered as the causal influence one time series variable has on another, i.e., bidirectional causation, which can be examined by Convergence Cross Mapping (CCM) [12]. CCM is a causation measure first used for variables in ecosystems. CCM reconstructs the state space for each variable using D-dimensional embedding of τ-delayed coordinates (i.e., the shadow manifold M) by assuming that there are shadow variables driving the underlying dynamics of the observed variable. When two variables are from the same dynamical state space (manifold), they are causally coupled. Using CCM, when two agents X, Y (with shadow manifolds M_X, M_Y) are perfectly coordinated, the points nearby M_X will match points nearby on M_Y for every time step in the manifolds. CCM then uses a simplex projection [30] technique to measure causality by finding the nearest $D + 1$ neighbors of M_X and the distances to those neighbors as weights to compute a cross-mapping estimation in M_Y.

B. Coordination in Different Stress Levels

1) Stress Level: When agents in a team experience different levels of demand (stress), the agents may behave differently. The stress levels can be determined by demand or urgency levels of an operation, computation load, or cognitive load for human agents. We aim to study the emergent coordination between agents as a team by analyzing agent decision behaviors at different stress levels. In our task allocation problem formulation, the operation environment state is represented by a number p of tasks with maximum task demand level l, and the total demand level ranges from 0 to pl. We categorize stress levels based on the total demand level. For example, three equally divided demand level ranges $[0, pl/3)$, $[pl/3, 2pl/3)$, $[2pl/3, pl]$ can represent low, medium, and high stress levels respectively.

2) Coordination between Two Agents: CCM requires long time series for convergence. In multi-agent task allocation, the decision time series are often short (i.e., less than 30 sequential decisions), depending on the number of agents, tasks, and task demand levels. Multi-Spatial Convergence Cross Mapping (MSCCM) [31] utilizes dewdrop regression [34] to compute

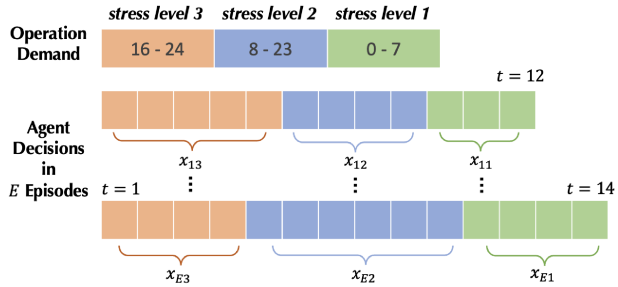


Fig. 1. Demonstration of aggregating short decision time series in $N = 3$ stress levels across E simulated episodes, for an operation with $p = 6$ tasks and $l = 4$ maximum task demand level.

CCM prediction skills with repeated observed experiences. Given the reproduciblility of a simulated environment with uncertainties, multiple episodes of short time series, sharing similar underlying dynamics, can be aggregated to create long time series. Therefore, we adapt MSCCM to discrete multi-agent Task Allocation CCM (TA-CCM) to quantify coordination at different stress levels by measuring the correspondence between the reconstructed state spaces (shadow manifolds) of two aggregated decision time series.

Let $i \in \{1, \ldots, E\}$ be the index of E simulated episodes of agents completing tasks, $j \in \{1, \ldots, N\}$ be the index of N stress levels, and let x_{ij}, y_{ij} denote the time series of task allocation decisions for agent x and agent y at stress level j in episode i. As demonstrated in Fig. 1, the aggregated decision time series for each stress level j are $X = [x_{1j}, \ldots, x_{Ej}]$, $Y = [y_{1j}, \ldots, y_{Ej}]$. Since each agent's decisions are segmented by stress level in the same way for each episode, X and Y have the same length. The shadow manifolds M_X, M_Y are constructed with lag τ and embedding dimension D, where $M_X(t) = [X_{(t)}, X_{(t-\tau)}, \ldots, X_{(t-(D-1)\tau)}]$ and $M_Y(t) = [Y_{(t)}, Y_{(t-\tau)}, \ldots, Y_{(t-(D-1)\tau)}]$. The simplex projection technique estimates the decisions of both agents using the other's shadow manifold for each time step t, denoted as $\hat{X}(t)|M_Y$ and $\hat{Y}(t)|M_X$, and the prediction skills are computed using the Pearson correlation coefficient ρ between the estimated and observed decisions, i.e. $\rho(\hat{X}|M_Y, X)$ and $\rho(\hat{Y}|M_X, Y)$ respectively. Such high prediction skills in an agent pair infer high coordination of the two agents.

C. Sub-Team Identification

With the help of TA-CCM described in Section III-B, the coordination in any agent pair can be measured using bidirectional prediction skills. Note that the complexity to compute the team coordination is $O(N^2 - N)$, and it can be reduced only if agents are known to be deterministic and predictable. One simple way of quantifying team-level coordination is to sum up the prediction skills between all pairs of agents. This approach requires high prediction skills between all agents to achieve high team-level coordination. One reasonable assumption is that the team has high coordination as long as every agent is highly coordinated with at least one other agent, and therefore multiple coordinated sub-teams with low prediction skills across different

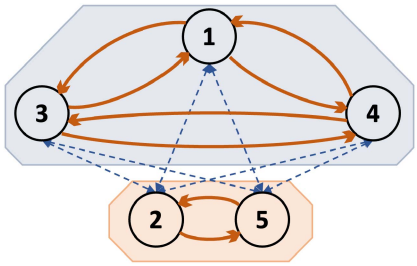


Fig. 2. A weighted directed network of a team of five agents with edges represented by behavior prediction skills. Solid orange arrows indicate strong predictability given an agent, dashed blue bidirectional arrows indicate weak predictability between agents, and shades indicate the two sub-teams.

sub-teams can have high team-level coordination. For example, as illustrated in Fig. 2, a team of 5 agents split into two sub-teams with strong predictability skills within the sub-teams can still be considered highly coordinated. Therefore, we construct a weighted directed network with vertices being agents and edges being TA-CCM prediction skills, and adapt a community finding technique: Modularity Maximization (ModMax) [13] for directed networks to identify statistically significant sub-teams. The number of sub-teams and the sub-team sizes do not need to be specified using ModMax.

Let $\mathcal{P}$ be edges of a directed network or an n by n matrix constructed by the prediction skills measured by TA-CCM, where n is the team size and $\mathcal{P}_{ij} = \rho(\hat{Y}_j|M_{X_i}, Y_j)$ denote the correlation coefficient between predicted decisions of agent j given agent i and observed decisions of agent j. Let κ_j^{in}, κ_i^{out} denote the in- and out-degrees of the agents, indicating how predictable agent j is and how useful agent i is to predict others respectively. We have $m = \sum_{ij} \mathcal{P}_{ij} = \sum_j \kappa_j^{in} = \sum_i \kappa_i^{out}$. ModMax allows us to find a division of a network with high benefit modularity U defined as:

$$U = \frac{1}{m} \sum_{ij} \left[\mathcal{P}_{ij} - \frac{\kappa_j^{in} \kappa_i^{out}}{m} \right] \delta_{g_i, g_j}, \tag{1}$$

where δ is the Kronecker delta, and g_i is the sub-team to which agent i is assigned. Large positive modularity U indicates statistically higher coordination within the sub-teams than expected. This process outputs a set of sub-teams $\{g\}$ that maximizes U given the prediction skill matrix $\mathcal{P}$.

D. Team Coordination

The sub-team identification process described in Section III-C tends to find sub-teams in which agents are more coordinated than those in different sub-teams. Therefore, we neglect the coordination across other sub-teams, and the team coordination only depends on the coordination levels and the sizes of the sub-teams. The team coordination score $\mathcal{I}$ can then be computed as:

$$\mathcal{I} = \sum_{\forall g} \omega_g C_g, \tag{2}$$

$$C_g = \frac{1}{|g|^2 - |g|} \sum_{\forall i \neq j} \mathcal{P}_{ij} \delta_{g_i, g_j}, \tag{3}$$

$$\omega_g = \frac{|g|}{n}, \tag{4}$$

where $\{g\}$ is the set of identified sub-teams, C_g is the sub-team coordination score, and ω_g is the sub-team size weight with respect to the full team size n. The sub-team coordination score C_g is the average prediction skills between different agents within the same sub-team. $|g|$ denotes the size of the sub-team, and $|g|^2 - |g|$ is the total number of directional prediction skills with self-prediction excluded. The team coordination metric is in the range of [0,1]. Given a prediction skill matrix with off-diagonal elements being 0 (i.e., everyone is unpredictable), the team would have 0 coordination score. When all sub-teams are perfectly coordinated ($C_g = 1$), the team coordination score will sum up to 1. The advantage of this metric based on sub-team identification and weighted by sub-team size is that it provides a generalized measure to all types of teams with different team sizes. Note that we can introduce a hyperparameter ψ to the sub-team size weight (i.e., $w_g = (|g|/n)^{\psi}$) to favor the coordinated sub-team with larger team size. Here, we choose $\psi = 1$ by assuming each coordinated sub-team contributes equally to the team coordination.

IV. Results

To analyze the emergent team coordination in multi-agent task allocation, we consider a Heterogeneous Teaming Decentralized Partially Observable Markov Decision Process (HT Dec-POMDP) [6] formulation, where the dynamics between task demand levels and agent capability levels are explicitly modelled, agents have their own decentralized decision models based on the beliefs over tasks, and each agent updates its belief by observing local demand and sharing its own observation and is rewarded upon the completion of tasks. Although the HT Dec-POMDP allows modeling heterogeneous tasks and agents where agents may specialize in certain tasks, it is more challenging to identify how a certain behavior emerges when tasks and agents are homogeneous, i.e., a symmetry breaking phenomenon [35]. In this study, we consider a number of indexed homogeneous tasks (i.e., the dynamics of all task demands are the same) and indexed agents with specialized task-specific capability. The demand level for each task is discrete and within the range of $[0, 4]$. The operation is complete when the demand levels of all tasks reach 0. The capability level of the agent task-specific attribute is 2 for the homogeneous team and can be either $1, 2,$ or 3 for the semi-homogeneous team (with a large number of agents, multiple agents have the same capability level). When two or more agents decide to collaborate on the same task, the joint capability level of the attribute is the sum of all agent capabilities, and capped at level 5. Higher capability level reduces the task demand level by 1 with higher probabilities. If the capability level is lower than the task demand level, the task demand level will increase, making the operation extraordinarily hard and requiring synchronized team coordination. It is also assumed that: 1) agents can accurately observe the task level at the assigned task with 80% probability, 2) agents are able to communicate their local observations with each other, and 3) the communicated observation could be either unaccepted,

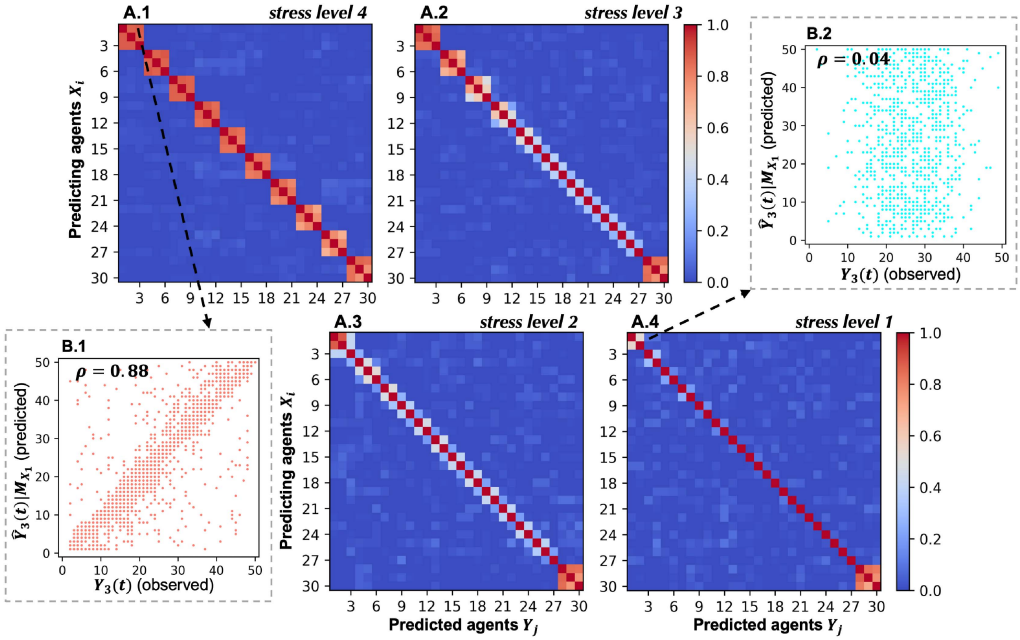


Fig. 3. (Panels A.1-4) The coordination for 4 stress levels for a team of 30 homogeneous agents in an operation with 50 homogeneous tasks, following cooperation and patrolling strategies. The coordination score $\rho \in [0, 1]$ between two agents is measured by TA-CCM with 1000 episodes. (Panels B.1-2) High and low correlations between observed and predicted decisions of agent 3 given agent 1.

wrongfully accepted, or accurately accepted for belief update and decision-making.

A. Decision Strategies

To demonstrate emergent team behaviors and evaluate team coordination, we consider explainable rule-based strategies for teams of homogeneous agents and one RL-trained policy described as the following:

Rule-Based - Agents have accurate full state observations. Lower indexed agents have priority in executing tasks with higher demand level.

- *Cooperation:* Homogeneous agents form groups of 3 to execute demand level 4 tasks, form groups of 2 to execute tasks with demand level 2 or 3, and individually execute demand level 1 tasks. Agents randomly explore the environment if there are no tasks available.
- *Patrolling:* A group of three homogeneous agents randomly picks three consecutive indexed tasks to execute.

RL-Trained - Agents make decentralized decisions and have partial observations. Agents are trained using Decentralized Deep Q-Learning with Beliefs [6] and make task allocation decision based on their own past experience and communicated information. Agents are trained to maximize the cumulative discounted reward for completing all tasks with fewer steps. During training, there are no predetermined rules on which agent should execute which task and no restrictions or penalty for switching tasks.

B. Coordination and Sub-Teams in Rule-Based Strategies

We first investigate the team coordinated behaviors given predefined rule-based strategies. We consider an operation with 50 homogeneous tasks with starting demand level at 4 and a team of 30 homogeneous agents with capability level at 2. The first 27 agents follow the cooperation strategy and the last three indexed agents follow the patrolling strategy as described in Section IV-A. The maximum possible operation demand level the team can experience is 200 (= level 4 $\times$ 50 tasks), and we equally categorize the operation demand levels into 4 ranges $[0, 50], [51, 100], [101, 150], [151, 200]$ that correspond to stress level $1 - 4$ respectively. To collect the task allocation decision time series data, agents play 1000 simulated episodes with up to 50 steps. Data collection stops when all tasks are completed. Decisions at each stress level are aggregated to create long decision time series.

Panels A.1-4 in Fig. 3 show the TA-CCM prediction skills between each pairs of agents at 4 stress levels. The diagonal entries are always 1 since these are self-predicted scores. By following the defined rule-based strategies, it is known that low indexed agents choose high demand level tasks. When multiple tasks with the same demand level are available, agent groups would choose the tasks randomly. Thus, it is hard to predict agent behaviors across different groups, resulting in low prediction skills in other entries. The last 3 agents always follow the patrolling strategy. Although the patrolling group executes three different consecutive tasks, TA-CCM is able to capture the strong correlations between these three agents. The rest of the agents follow the cooperation strategy. At stress level 4, most of the tasks have demand level 4, and agents almost always form groups of three. At stress levels 3 and 2, tasks have various demand levels, and agents can form groups of any size up to 3. Since lower indexed agents are preferred to execute high demand level tasks, it is more likely to see high correlation groups (top left corner) in lower indexed agents. At

TABLE I
TEAM SPECIFICATIONS

Team	Decision Strategy	Uncertainties in Decision/Observation	Steps to Complete All Tasks
1	Rule-based	80% rational	31.6
2	RL-trained	80% accurate	26.4
3	RL-trained	80% accurate	25.7

the lowest stress level, most of the tasks are completed, and agents depict random behaviors and low correlations with each other.

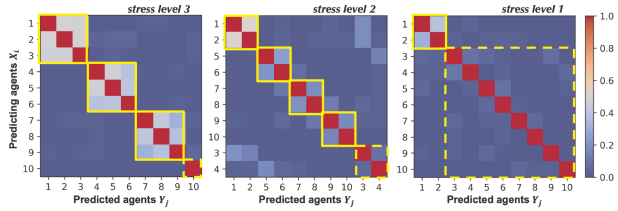
Panels B.1-2 in Fig. 3 illustrate the prediction skills measured by the correlation ρ between the observed decisions Y_3 and the predicted decisions of agent 3 given agent 1's shadow manifold $\hat{Y}_3|M_{X_1}$ in stress level 4 and 1. Note that TA-CCM is not used for the purpose of predicting the exact decisions of agent 3, it measures the trend and correlation between two agent behaviors. We can conclude that agent 3 behavior can be inferred by using agent 1 behavior more easily, and therefore more coordinated at stress level 4 than stress level 1, because agent 1 and 3 are more likely in the same group to execute high demand level tasks and are in different groups to execute low demand level tasks.

C. Coordination and Sub-Teams in Learned Strategy

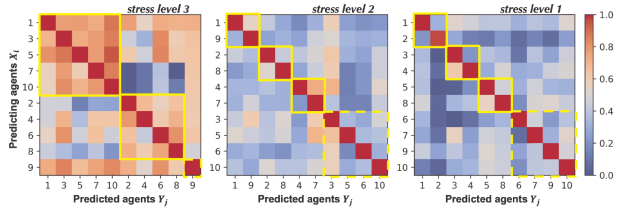
In predefined rule-based strategies, we design a priori the agent groups under different task demand conditions. Coordinated behaviors are clearly shown in the prediction skill matrices. However, with only goal-oriented reinforcement learning agents, it is ambiguous (before training) which agents have statistically significant coordinated decision behaviors and thus can form a sub-team. In this section, we apply TA-CCM and ModMax methods to identify the emergent sub-teams and measure the overall team coordination score for teams of 10 RL-trained decentralized agents in an operation with 14 homogeneous tasks. We consider only 10 agents to reduce computation time of training, and 14 tasks are challenging enough to create long decision time series.

We consider 3 different teams and collect their decisions with uncertainties in 2000 episodes. The performance of each team is the average number of steps taken to reduce all task demand levels to 0. The team specifications are summarized in Table I. Team 1 and 2 both consist of 10 homogeneous agents with capability level 2. Team 3 is semi-homogeneous and has 10 indexed agents with capability levels [3,3,3,2,2,2,2,1,1,1]. Since cooperation grants Team 1 with full state observation, we add uncertainties in making rational decisions with 20% probability of taking random actions, while RL-trained agents receive accurate local observations with 80% probability.

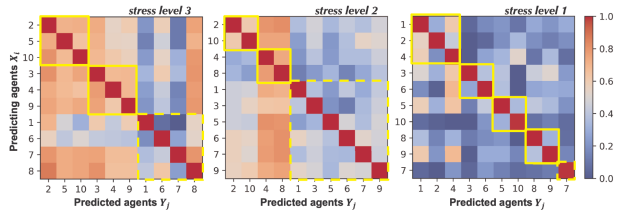
Fig. 4 shows the prediction skills measured by TA-CCM and sub-teams identified by ModMax for 3 teams. The agent indices are reorganized by the sub-teams. Having 14 tasks with demand level 4, the 3 stress levels correspond to 3 divided task demand level ranges $[0, 18]$, $[19, 37]$, $[38, 56]$. ModMax is able to identify the sub-teams as the predefined groups in cooperation strategy. Since ModMax does not have restrictions on the size and number



(a) Homogeneous agents with Rule-based policy



(b) Homogeneous agents with RL-trained policy



(c) Semi-Homogeneous agents with RL-trained policy

Fig. 4. Sub-team identification for three 10-agent teams under cooperation rules or RL-trained policy in 3 stress levels. Solid yellow boxes indicate identified sub-teams and dashed yellow boxes show ungrouped agents.

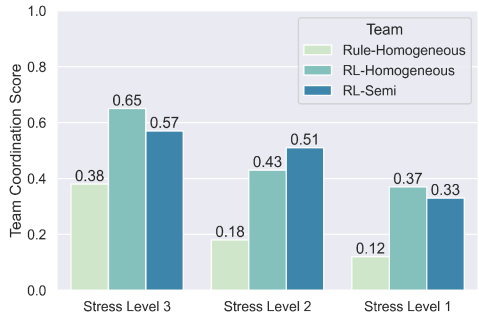
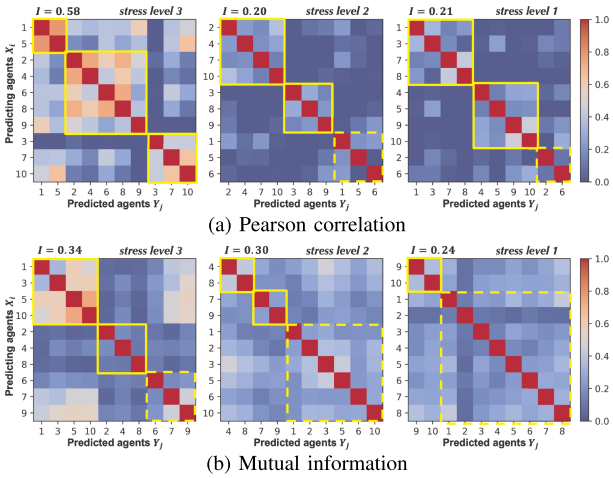


Fig. 5. Coordination scores for three teams at different stress levels.

of sub-teams, it is able to identify larger sub-teams in the RL-trained homogeneous team. Note that we exclude agents that do not belong to any sub-teams, since ungrouped agents have statistically insignificant contribution to team coordination.

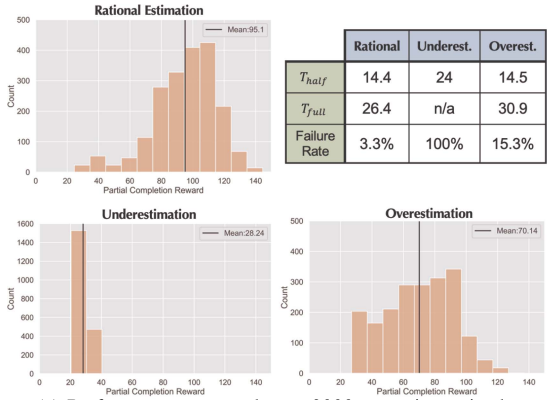
Then, we apply the team coordination metric described in III-D to evaluate the synchronization of three teams. As shown in Fig. 5, RL-trained teams illustrate higher coordination than the cooperation team facing an equivalent level of uncertainties, and all teams demonstrate higher coordination in higher stress level situations. Since the cooperation strategy only reacts to the current task situation, it only captures spatial coordination (i.e., agents execute the same task). In contrast, the RL-trained teams maximize the reward for a sequence of actions, and are able to capture both temporal (i.e., action sequence) and spatial coordination and consequently are more coordinated. The high coordination score at high stress levels explains the need for temporal and spatial coordination under severe situations. The



(a) Pearson correlation

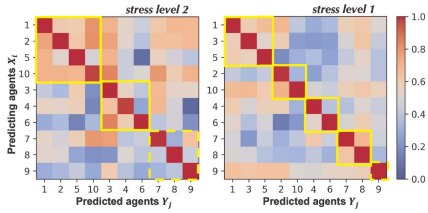
(b) Mutual information

Fig. 6. Sub-team identification and coordination scores for the 10-agent homogeneous team under RL-trained policy, using (a) Pearson correlation and (b) mutual information.



	Rational	Underest.	Overest.
T_{half}	14.4	24	14.5
T_{full}	26.4	n/a	30.9
Failure Rate	3.3%	100%	15.3%

(a) Performance averaged over 2000 operation episodes.



(b) Sub-team identification when overestimating. Team coordination score: 0.56 and 0.50 at stress level 2 and 1 respectively.

Fig. 7. Performance of the team of 10 RL-trained homogeneous agents when making rational decisions, compared to overestimating and underestimating the stress level.

coordination scores at lower stress levels imply uncertainty in environment demands and stochasticity of team strategy.

To verify the interpretability of TA-CCM, we illustrate the coordination evaluation in Fig. 6, using Pearson correlation and mutual information as coordination measures of each agent pair. Both metrics show symmetric coordination between agent pairs, which neglect the causation effect in agent decisions and therefore reach lower team coordination scores, compared to the RL-Homogeneous Team using TA-CCM. The Pearson correlation can only identify high coordination in stress level 3 because agents tend to work on the same task at high stress levels, but the estimated coordination across different sub-teams and in lower stress levels is much lower than TA-CCM. While mutual information measures how much knowing one variable can reduce the uncertainty of the other, the distribution of an agent's decision does not contain temporal information, and consequently mutual information fails to identify coordinated sub-teams. Both measures identify different sub-teams and weaker coordination in large sub-teams, compared to TA-CCM.

D. Underestimation and Overestimation

In practice, agents do not always accurately assess the demands and the stress level of the situation, and the dynamics of demands might change due to unexpected events such as degraded agent capability or sudden demand increase. The changes in team behaviors need to be recognized and adjusted accordingly to avoid catastrophic failures. In this study, considering the same operation as in Section IV-C, we utilize TA-CCM and ModMax to analyze unexpected coordination strategies when the RL-trained homogeneous team under-/over-estimates the environment task demands. To exaggerate the effect of under-/over-estimation, all agents start to believe that the task demand level is halved/doubled once the environment reaches stress level 2. For example, in underestimation, the agent believes the task demand level is 2 while the actual level is 3 or 4; in overestimation, the agent believes the task demand level is 4 while the actual level is 2. The agent would then make decisions based on the under-/over-estimated demand levels. We consider four performance metrics for under-/over-estimating and making rational decision (accurate estimation) averaged over 2000 episodes as shown in Fig. 7(a): 1) T_{half}: number of steps to complete half of tasks, 2) T_{full}: number of steps to complete all tasks, 3) Failure Rate: chances of not completing all tasks within 50 steps over 2000 episodes, and 4) Partial Completion Reward: cumulative discounted reward for completing a portion of tasks (i.e., bringing half of the tasks to level 0 receives half of the completion reward), incentivizing completing tasks in earlier steps. When agents underestimate tasks, they are never able to complete all the tasks. When agents overestimate the situation, they manage to complete all tasks but not as efficiently as making optimal decisions. This is expected since overestimation requires agents to execute the same tasks more often while they could have handled multiple tasks simultaneously. The performance of overestimation is verified through the sub-team identification in Fig. 7(b). At stress level 1 and 2, the RL-trained homogeneous team demonstrate similar behaviors and coordination scores as at stress level 3 (Fig. 4(b)). Deploying different coordination strategies rather than the trained strategy at different stress levels might result in reduced in performance, despite of high coordination score.

V. Conclusion

In this letter, we developed a metric to quantify team coordination in different task demand (stress) levels. The coordination measure was applied to multiple team compositions under various coordination strategies. The approach detects the causality

effect between agents in a pair using decision time series data, and identifies the highly coordinated sub-teams within the multi-agent teams using a community finding technique in weighted networks. Results of the study show that the proposed measure can identify higher emergent coordination level for RL-trained strategies across all stages of an operation, and demonstrate that applying a proper coordination pattern at different stages of an operation leads to successful teaming without catastrophic failures.

Understanding the strategies of centralized rules and the evolved strategies of incentive-driven reinforcement learning is fundamental for further improvement in creating large-scale artificial intelligence multi-agent systems, especially in cases where adaptive strategies are needed to ensure the robustness and the efficiency of team coordination under unforeseen situations. The coordination metric along with the identified sub-teams can provide the following benefits in evaluating rule-based strategies and training multi-agent RL agents: 1) detecting whether the learned strategy and sub-team formation are proper in corresponding situations, 2) monitoring whether the team and the sub-teams can maintain a high level coordination when situations become unpredictable, 3) potentially simplifying team design with sub-teams, and 4) improving the coordinated learning by using coordination score as a guidance. The proposed method provides opportunities to acknowledge how communities (high coordinated sub-teams) or leader-follower (one directional causality) relations are forming and influencing the coordination strategies in various types of multi-agent teams.

REFERENCES

[1] Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," *IEEE Trans. Ind. Inform.*, vol. 9, no. 1, pp. 427–438, Feb. 2013.

[2] S. Gronauer and K. Dieopold, "Multi-agent deep reinforcement learning: A survey," *Artif. Intell. Rev.*, vol. 55, pp. 1–49, 2022.

[3] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2008, pp. 1433–1438.

[4] D. Ramachandran and E. Amir, Burlington, MA, USA: Morgan Kaufmann Publishers, 2007, pp. 2586–2591.

[5] S. Natarajan, G. Kunapuli, K. Judah, P. Tadepalli, K. Kersting, and J. Shavlik, "Multi-agent inverse reinforcement learning," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, 2010, pp. 395–400.

[6] H. Wu, A. Ghadami, A. E. Bayrak, J. M. Smereka, and B. I. Epureanu, "Impact of heterogeneity and risk aversion on task allocation in multi-agent teams," *IEEE Robot. Automat. Lett.*, vol. 6, no. 4, pp. 7065–7072, Oct. 2021.

[7] H. Wu, A. Ghadami, A. E. Bayrak, J. M. Smereka, and B. I. Epureanu, "Task allocation with load management in multi-agent teams," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2022, pp. 8823–8830.

[8] Y. Zhang, Q. Yang, D. An, and C. Zhang, "Coordination between individual agents in multi-agent reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 11 387–11 394.

[9] S. Liu, G. Lever, N. Heess, J. Merel, S. Tunyasuvunakool, and T. Graepel, "Emergent coordination through competition," in *Proc. 7th Int. Conf. Learn. Representations*, New Orleans, LA, USA, May 6-9, 2019. [Online]. Available: https://openreview.net/forum?id=BkG8sjR5Km

[10] S. Barton, N. Waytowich, E. Zaroukian, and D. Asher, "Measuring collaborative emergent behavior in multi-agent reinforcement learning," in *Proc. 1st Int. Conf. Hum. Syst. Eng. Des.*, 2018, pp. 422–427.

[11] D. Asher, S. Barton, E. Zaroukian, and N. Waytowich, "Effect of cooperative team size on coordination in adaptive multi-agent systems," *Artif. Intell. Mach. Learn. Multi-Domain Operations Appl.*, vol. 11006, pp. 337–348, 2019.

[12] G. Sugihara et al., "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.

[13] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Phys. Rev. Lett.*, vol. 100, 2008, Art. no. 118703.

[14] C. Guestrin, M. G. Lagoudakis, and R. Parr, "Coordinated reinforcement learning," in *Proc. 19th Int. Conf. Mach. Learn.*, 2002, pp. 227–234 .

[15] R. Nair, P. Varakantham, M. Tambe, and M. Yokoo, "Networked distributed POMDPs: A synthesis of distributed constraint optimization and POMDPs," in *Proc. Proc. 20th Nat. Conf. Artif. Intell.*, 2005, vol. 1, pp. 133–139.

[16] W. Böhmer, V. Kurin, and S. Whiteson, "Deep coordination graphs," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 980–991.

[17] E. Puiutta and E. M. S. P. Veith, "Explainable reinforcement learning: A survey," in *Proc. Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction*, 2020, pp. 77–95.

[18] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, pp. 206–215, 2019.

[19] T. Zahavy, N. Ben-Zrihem, and S. Mannor, "Graying the black box: Understanding DQNs," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, vol. 48, pp. 1899–1908.

[20] P. Sequeira and M. Gervasio, "Interestingness elements for explainable reinforcement learning: Understanding agents' capabilities and limitations," *Artif. Intell.*, vol. 288, 2020, Art. no. 103367.

[21] E. A. O. Diallo and T. Sugawara, "Learning strategic group formation for coordinated behavior in adversarial multi-agent with double dqn," in *Proc. Princ. Pract. Multi-Agent Syst.*, 2018, pp. 458–466.

[22] Y. Miyashita and T. Sugawara, "Analysis of coordinated behavior structures with multi-agent deep reinforcement learning," *Appl. Intell.*, vol. 51, no. 2, pp. 1069–1085, 2021.

[23] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proc. Roy. Soc. London Ser. I*, vol. 58, pp. 240–242, 1895.

[24] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.

[25] P. N. Sabes and M. Jordan, "Reinforcement learning by probability matching," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 1995, pp. 1080–1086.

[26] E. Zaroukian et al., "Algorithmically identifying strategies in multi-agent game-theoretic environments," *Artif. Intell. Mach. Learn. Multi-Domain Operations Appl.*, vol. 11006, 2019, Art. no. 1100614.

[27] C. W. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica: J. Econometric Soc.*, vol. 37, pp. 424–438, 1969.

[28] H. Ye, E. Deyle, L. Gilarranz, and G. Sugihara, "Distinguishing time-delayed causal interactions using convergent cross mapping," *Sci. Rep.*, vol. 5, 2015, Art. no. 14750.

[29] L. Breston, E. Leonardis, L. Quinn, M. Tolston, J. Wiles, and A. Chiba, "Convergent cross sorting for estimating dynamic coupling," *Sci. Rep.*, vol. 11, 2021, Art. no. 20374.

[30] J. M. McCracken and R. S. Weigel, "Convergent cross-mapping and pairwise asymmetric inference," *Phys. Rev. E*, vol. 90, 2014, Art. no. 0 62903.

[31] A. T. Clark et al., "Spatial convergent cross mapping to detect causal relationships from short time series," *Ecol.*, vol. 96, no. 5, pp. 1174–1181, 2015.

[32] M. E. J. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, 2004, Art. no. 0 56131.

[33] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein, "The complexity of decentralized control of Markov decision processes," vol. 27, no. 4, pp. 819–840, 2002.

[34] C. Hsieh, C. Anderson, and G. Sugihara, "Extending nonlinear analysis to short ecological time series.," *Amer. Naturalist*, vol. 171, no. 1, pp. 71–80, 2008.

[35] H. Hamann, T. Schmickl, H. Wörn, and K. Crailsheim, "Analysis of emergent symmetry breaking in collective decision making," *Neural Comput. Appl.*, vol. 21, pp. 207–218, 2012.