

Automated Task-Time Interventions to Improve Teamwork using Imitation Learning

Sangwon Seo
Rice University
Houston, TX, United States
Sangwon.Seo@rice.edu

Bing Han
Rice University
Houston, TX, United States
bh39@rice.edu

Vaibhav Unhelkar
Rice University
Houston, TX, United States
Vaibhav.Unhelkar@rice.edu

ABSTRACT

Effective human-human and human-autonomy teamwork is critical but often challenging to perfect. The challenge is particularly relevant in time-critical domains, such as healthcare and disaster response, where the time pressures can make coordination increasingly difficult to achieve and the consequences of imperfect coordination can be severe. To improve teamwork in these and other domains, we present TIC: an automated intervention approach for improving coordination between team members. Using BTIL, a multi-agent imitation learning algorithm, our approach first learns a generative model of team behavior from past task execution data. Next, it utilizes the learned generative model and team’s task objective (shared reward) to algorithmically generate execution-time interventions. We evaluate our approach in synthetic multi-agent teaming scenarios, where team members make decentralized decisions without full observability of the environment. The experiments demonstrate that the automated interventions can successfully improve team performance and shed light on the design of autonomous agents for improving teamwork.

CCS CONCEPTS

• **Computing methodologies** → **Intelligent agents**; *Planning under uncertainty*; *Multi-agent planning*; *Machine learning*; • **Human-centered computing**;

KEYWORDS

Teamwork; Markov Decision Processes; Imitation Learning; Bayesian Inference; Computer-Aided Human Decision-Making

ACM Reference Format:

Sangwon Seo, Bing Han, and Vaibhav Unhelkar. 2023. Automated Task-Time Interventions to Improve Teamwork using Imitation Learning. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023)*, London, United Kingdom, May 29 – June 2, 2023, IFAAMAS, 10 pages.

1 INTRODUCTION

The success of human enterprise depends on effective teamwork. In domains as diverse as manufacturing, disaster response, and healthcare rarely a single agent conducts a task alone; instead, to benefit from complementary expertise of multiple and diverse agents, processes are almost always multi-agent and collaborative. In certain situations the teaming context is well defined, while in others the

teamwork might be loosely-coupled [2, 10, 23, 31]. Further, real-world teams may be composed of humans alone or, with ongoing advances in artificial intelligence, include robots and autonomous agents. Irrespective of the teaming context or composition, teamwork is often challenging to perfect. Imperfect coordination among team members can result in unsatisfactory performance or failures in collaborative tasks [21, 50]. The effect of these imperfections can be particularly severe in time- and life-critical domains, such as disaster response and healthcare, where the time pressures can make coordination increasingly difficult to achieve.

Recognizing the ubiquity and adverse consequences of imperfect coordination, there is a long and growing body of research on studying and improving teamwork in multiple disciplines. For instance researchers in team science, by adopting a human factors perspective, have developed theory and methods for team training and assessments [4, 6, 25, 29, 35, 36, 48]. Concurrently, artificial intelligence community has developed models and algorithms for perfecting multi-agent systems [10, 28, 31, 44]. While methods and techniques differ across research communities, not least due to the nuances of different teaming contexts, certain shared insights for improving teamwork exist. For instance, team members must maintain situational awareness of the task context and mental models of their team members and, by leveraging this situational awareness, adapt their plans to achieve and maintain coordination.

In practice, however, realizing this insight is challenging for team members without appropriate interventions. In many collaborative contexts, agents need to act under partial observability of the task context as well as that of belief, desires, and intentions of their team members [31]. This partial observability makes it challenging for team members to maintain situation awareness and correctly adapt to maintain shared plans. For this reason, in some domains such as team sports, a team employs a coach who has a better sight of the task environment and can intervene to maintain team coordination [3, 46]. In most other domains, however, resource constraints make it difficult to find an expert who can oversee a team and intervene during each task execution. Instead, teams adopt team science-inspired methods to improve coordination. For instance, successful teams participate in training and debriefing sessions [29, 36, 48]. They may also conduct self-diagnostic surveys to assess their teamwork [4, 6, 25]. However, these methods are either *ex ante* or *post hoc* and cannot provide task-time interventions.

To complement these methods, in this work, we consider the *problem of algorithmically generating task-time interventions to improve teamwork*. As reviewed in Sec. 2, there is growing interest in developing artificial agents to improve human decision-making.

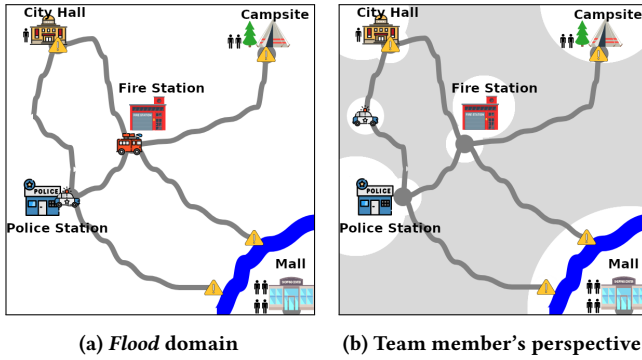


Figure 1: A domain inspired by disaster response, where a team has to rescue people in a decentralized and partially observable environment.

However, to our knowledge, this work is the first to explore feasibility of algorithmically generated *task-time* interventions to improve teamwork. To solve this problem, we provide the algorithm Task-time Interventions for Improving Collaboration (TIC) that builds upon recent advances in multi-agent imitation learning [39] and, given a Markovian model of the collaborative task [31] and demonstrations, generates automated task-time interventions.

1.1 Running Example

To illustrate our problem effectively, we consider a scenario inspired by disaster response [18]. Fig. 1a depicts the scenario, where a team of two members – a police officer and a firefighter – is requested to rescue people collaboratively after a flood. As further shown in Fig. 1b (unshaded area), the team members have to operate under partial observability; in particular, each team member can observe their teammate only in their immediate vicinity or at landmarks. There are four locations (marked with a yellow sign) where people are stranded: a city hall, a campsite, and two broken bridges, each with a different number of people. While people at city hall and campsite can be rescued just by one agent, repairing a bridge requires both the police officer and firefighter to work together. As people in the mall can be rescued by repairing just one of the bridges, the team does not have to repair all the bridges.

Each agent can take one of the following actions: move to one of the directions if they are on the road, stay put, or rescue people if they are at a disaster location (yellow sign). Due to the time-critical nature of the task, in this running example, the team does not have time for prior joint planning and needs to coordinate on-the-fly. To complete the task effectively (i.e., rescue the most number of people within the time limit), the team members need to coordinate the sites to visit and select cooperative action to realize the shared objective. As such, the behavior of a coordinated team must depend both on the task context (e.g., location of landmarks, state of the bridge) and team member’s mental states (e.g., belief about which sub-activity to pursue next, intent regarding next landmark to visit).

1.2 Overview of the Proposed Approach

As coordination is both critical and challenging in this and other time-critical collaborative tasks, TIC adopts a data-driven approach

to generate immediate interventions to improve team coordination during the task execution. TIC involves modules to

- C1. **model** each member’s behavior and mental states,
- C2. **detect** when teamwork is poorly coordinated, and
- C3. **intervene** as appropriate to improve the teamwork.

To realize capability C1, TIC requires a behavioral model of each team member that depends on both the task context and team members’ mental states (e.g., belief, sub-goals, or intent). For instance, in our running example, C1 corresponds to prediction of team member’s intended sub-goals (e.g., which bridge to repair next) and goal-dependent policies. Planning or rule-based approaches have been used in the past to model team behavior; however, these techniques either do not account for team member’s mental states or, in realistic settings, require significant manual effort to encode behavior for each context [1]. Recently, imitation learning has garnered popularity in learning predictive models of agent behavior from data [22, 32, 34], with growing interest in multi-agent imitation learning [20, 42, 49]. Informed by successes of these works, we utilize Bayesian Team Imitation Learning (BTIL), a recent multi-agent imitation learning technique to realize C1 [38]. Our choice of the imitation learning algorithm is informed by BTIL’s ability to learn behavioral models of teamwork that (a) explicitly depend on team members’ mental states, and (b) model team behavior under both perfect and imperfect coordination settings.

During execution of collaborative tasks, poor coordination can result from a variety of factors: misalignment in team member’s mental states, imperfect task execution by a particular team member, among others. For instance, in our running example, the team members may each select a different bridge to repair resulting in imperfect coordination, lost time, and poor task outcomes. In this work, we focus on imperfect coordination resulting from team member’s mental states, which has been identified as a significant factor influencing teamwork [5, 35]. As team member’s mental states are not readily observable, solutions are needed to infer them from observed data. Thus, to realize capability C2 in TIC, we propose a Bayesian approach that estimates at *task time* alignment or lack thereof in team member’s mental states. The proposed Bayesian approach, described in Sec. 4.3, utilizes the team behavioral model learned using C1 and the observed context (state) of the collaborative task as inputs for mental state inference.

Interventions, when appropriate, can improve teamwork. For instance, in our running example, interventions that help rectify misalignment in team member’s mental states (e.g., which bridge to repair next) can improve task outcomes. At the same time, however, too many interventions can distract teams from their tasks and lead to unintended side effects (e.g., switching costs in human teams) [14]. Thus, approaches that can effectively trade-off cost and benefit of interventions are needed. In Sec. 4.5, we propose and study multiple intervention strategies to realize C3 in TIC. The proposed strategies determine when to intervene based on the team’s objective, team behavioral model learned via C1, and the estimate derived using C2. We emphasize development of strategies that generate context-specific interventions and require minimal manual input to estimate the utility of intervention.

We evaluate TIC with four synthetic teaming scenarios, each inspired by real-world applications and existing research [18, 39].

The experiments demonstrate that the proposed approach TIC can generate automated interventions to improve team performance and can effectively trade-off the cost and benefit of interventions. These results help establish the feasibility of automated task-time interventions. Further, the empirical analysis comparing different intervention strategies (C3) of TIC provide guidance for the design of future AI systems to improve human teamwork.

2 RELATED WORK

Our work is informed by a rich body of research on teamwork, both in humanities and artificial intelligence [4, 6, 10, 25, 28, 29, 31, 35, 36, 44, 48]. In this brief review, we discuss approaches for team assessment (relevant to C2) and intervention (relevant to C3).

Methods for Team Assessment. Researchers in team science have established robust practices for assessing human-only teams [8, 19, 35]; however, these assessment approaches are post-hoc and do not involve automation. As AI agents increasingly become part of human teams, recent studies in human-robot interaction suggest several metrics for the teamwork of various types of teams beyond human teams [12, 24, 30]. However, to our knowledge, no methods exist to compute these metrics at task time using automated techniques. Other works present team plan or team intention recognition methods to predict team behavior [27, 37, 51], but do not explicitly draw the connection between team intention and team effectiveness. Seo et al. [38] leverage the concept of shared mental models to recognize ineffective teamwork based on the recognition of each member’s mental state. However, they assume perfect knowledge of team behavior and do not explore the problem of when to intervene to achieve effective teamwork.

Enhancing Teamwork using AI. Across domains, there is growing interest in developing artificial agents to improve teamwork [2, 7, 15, 17, 18, 38, 46]; however, to our knowledge, our work is the first to explore automated *task-time* interventions in tightly-coupled and time-critical tasks. For instance, Amir et al. [2] present a model of information sharing between the loosely-coupled team, which helps decide when to share information between partners based on the concept of degree of interest. Kim and Shah [18] provide an approach to enhance teamwork by developing an intelligent agent to improve human team planning. Their approach detects the degree of shared understanding by analyzing the dialogues in a meeting before conducting the task. Tausczik and Pennebaker [45] provide real-time language feedback based on the communication patterns of a team. In practice, our approach to generating *task-time* interventions could serve as a useful complement to these existing approaches for generating *planning-time* interventions.

3 PROBLEM FORMULATION

To address the problem of generating automated task-time interventions for improving teamwork, we first formalize it using existing models of teamwork [31]. Further, we focus on collaborative tasks with a well-defined notion of shared objective. While our formalism is agnostic to team composition, it is particularly relevant for teams with no hierarchy and at least one human member, i.e., equal partners human teams or human-robot teams.

3.1 Model of Collaborative Task

We focus on teams that perform finite-horizon tasks with well-designed objectives in partially observable environments. To formulate team behavior while taking into account the relevant uncertainties, we model the task as a decentralized partially observable Markov decision process (Dec-POMDP). Dec-POMDP represents a task by a tuple $(n, S, A, \Omega, T_s, O, R, \gamma, h)$, where n is the number of team members, S is the set of task states, A is the set of joint actions a of team members, Ω is the set of joint observations of team members, $T_s : S \times A \times S \rightarrow [0, 1]$ denotes the state transition probabilities, $O : \Omega \times A \times S \rightarrow [0, 1]$ denotes the observation probabilities, $R : S \times A \rightarrow \mathbb{R}$ denotes the shared reward, γ is the discount factor, and h is the task horizon. The joint action $a \doteq (a_1, \dots, a_n)$ can be factorized as $A \doteq \times_i A_i$, where A_i denotes the set of the i -th member’s actions $a_i \in A_i$. For more details on Dec-POMDP, we refer the reader to the text by Oliehoek and Amato [31].

3.2 Model of Team Behavior

In theory, given the task model, a team can act optimally by computing a decentralized policy using Dec-POMDP solvers and then executing it in lockstep. However, such optimal behavior is unrealistic to expect in practice. First, computing optimal policies for Dec-POMDP models remains computationally prohibitive for many practical problems, due to which multi-robot teams often employ heuristic policies. Further, even when an optimal policy can be computed, it is unrealistic to expect team members (particularly humans as well as imperfect robots) to follow such a policy in lockstep and without any execution errors. In practice, behavior of human team members is often bounded rational and depends on mental states, such as personal preferences or physiological states [9, 11, 41]. To reflect the dependence of team member’s behavior on the mental states, we explicitly model task-relevant mental states as $x_i \in X_i$ and assume them to be unobservable for all team members $j \neq i$.

Borrowing from the Agent Markov Model [47], we represent the behavior of the i -th member to depend on the tuple $(X_i, b_{x_i}, T_{x_i}, \pi_i)$, where X_i is the set of their mental states, b_{x_i} is the probability distribution of the mental state at time 0, T_{x_i} denotes the transition dynamics of the mental state, and π_i denotes the member’s policy. Concretely, T_{x_i} denotes the probability of i -th team member’s mental state at the next step, conditioned on the task history and their current mental state. Similarly, π_i denotes their probability of selecting action $a_i \in A_i$ conditioned on the task history and their current mental state. While in general the tuple $(X_i, b_{x_i}, T_{x_i}, \pi_i)$ will be unknown for each team member, for tractability, we assume that the set of mental states X_i for each member is known and finite. Similar to joint actions, we denote $x \doteq (x_1, \dots, x_n)$, and $X \doteq \times_i X_i$.

3.3 Problem Statement

Having identified the models of task and team behavior, we next formalize the problem of interest. The space of potential interventions to improve teamwork is large. A coach may intervene a team to convey optimal actions, mental states, plans, among other variables of interest [46]. In this work, we limit the intervention content to recommended value of mental states (x) which improve team coordination. We leave the exploration of other types of intervention to future work and, instead, focus on the problem of deciding whether

or not to intervene to improve team performance. We use z to denote the binary decision of intervening in team’s task execution. The variable c denotes the cost per intervention and serves as a simple model for adverse consequences of intervention [14, 26]. Formally, the problem assumes as input

- the task model $(n, S, A, \Omega, T_s, O, R, \gamma, h)$,
- the set of team member’s mental states (X) , and
- task execution data $(s^{0:t}, a^{0:t-1})$ up to time t .

Given these inputs, the problem corresponds to deciding z_t (i.e., whether or not to intervene at time t) so as to maximize the objective $J = \sum_{t=0}^h (r_t - c_t z_t)$. Note that the problem assumes that, similar to a sports coach, the intervention system has full observability of the task state (s) but not that of team members’ mental states (x). The objective reflects the trade-off of maximizing team’s cumulative reward, while minimizing the cumulative cost of interventions.

4 SOLUTION

A trivial solution to this problem is to always intervene; however, this solution is likely to be sub-optimal due to the cost of intervention. Hence, we propose an approach that seeks to generate effective interventions by reasoning about team members’ behavior and mental states. In particular, to derive the solution for the formulated problem, we break down the problem into three sub-problems

- P1. **Model:** Given the task model $(n, S, A, \Omega, T_s, O, R, \gamma, h)$, the set of team member’s mental states (X) , and partially annotated sequences of team’s past m task executions $(s^{0:h}, a^{0:h}, x^{t < h})$, *learn a model of each team member’s behavior* $(b_{x_i}, T_{x_i}, \pi_i) \forall i$.
- P2. **Detect:** Given the inputs of P1, a learned model of each team member’s behavior $(b_{x_i}, T_{x_i}, \pi_i) \forall i$, and observable data of on-going task execution $(s^{0:t}, a^{0:t-1})$, *estimate at task-time each team member’s mental state* (x^t) .
- P3. **Intervene:** Given the inputs of P2 and an estimate of each team member’s mental state (x^t) , *decide at task-time whether to intervene* (z_t) so as to maximize the objective J .

Our approach includes three key modules, one corresponding to each sub-problem described above. Fig. 2 depicts a schematic of our approach. As motivated in Sec. 1.2, first, we use multi-agent imitation learning techniques to solve P1. Next, use the outcome of P1 and a Bayesian inference technique to solve P2. And, lastly, propose and study a set of intervention strategies to solve P3.

4.1 Multi-Agent Imitation Learning

To address P1, we observe that the sub-problem of modeling team behavior can be posed as one of imitation learning. Mathematically, the classical imitation learning setting seeks to learn a single-agent policy (π), a function that returns the probability of selecting an action (a) given a task context (s), from a dataset of task executions: (s, a) -pairs. Recently, there is growing interest in addressing the multi-agent version of this problem setting, where the goal is to learn decentralized policy ($\pi_i \forall i$) of a multi-agent system from demonstrations [20, 42, 49]. Most multi-agent imitation learning techniques, however, do not explicitly model agent’s mental states (X) or learn its temporal dynamics (T_x).

As motivated in Sec. 1.2, multi-agent behavior in teaming context depends not only on the task context (s) but also on a variety of

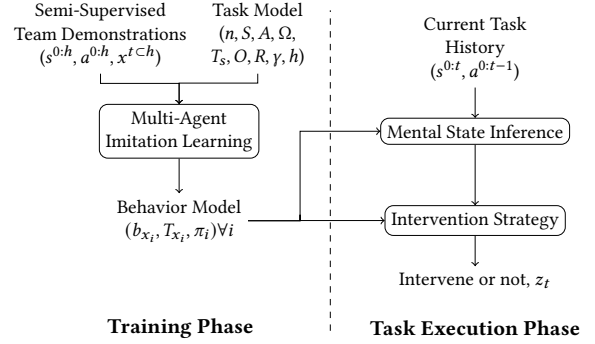


Figure 2: TIC schematic: During the training phase, TIC learns a behavioral model of the team from past execution data. Upon training, when a team performs a task, TIC detects poor coordination and intervenes to improve teamwork.

latent decision factors (x), such as preference, sub-goals, or belief. A few recent papers explicitly include latent factors as independent variables in the behavior model to better account for complex agent behavior [16, 39, 40, 47]. Among these, only the approach Bayesian Team Imitation Learning (BTIL) enables learning of team members’ decentralized policies and mental state dynamics from suboptimal demonstrations [39].

Briefly, BTIL assumes a Markovian model for the dynamics of team members’ mental states $T_{x_i}(x'_i | s, x_i, a, s')$ and policy $\pi_i(a_i | s, x_i)$, and that each agent can fully observe the environment. Under these assumptions, BTIL computes the maximum a posteriori (MAP) estimate of T_x and π from past task-execution data. Since the posterior of T_x and π cannot be computed analytically due to latent variables, BTIL utilizes mean-field variational inference (MFVI) [13]. With MFVI, BTIL approximates the posterior $p(T_x, \pi | \text{data})$ by maximizing the evidence lower bound (ELBO). For more details of BTIL, we refer the reader to [39].

4.2 Learning Models of Team Behavior

The problem setting of BTIL corresponds closely to that of P1. Both settings focus on decentralized team policies, consider mental states and their dynamics, and do not assume perfect coordination in the demonstrations of team behavior. However, certain key differences also exist. Most importantly, BTIL considers the setting where team members have full observability of the task state, while our problem setting considers partially observable tasks. As such, the behavior in BTIL is modeled using state-dependent policies and not using the general case of history-dependent policies. Learning history-dependent policies of single agents, let alone teams, is a challenging open problem [1].

However, as we seek to model team behavior from the intervention system’s point of view, we can leverage its full observability of the task state during the modeling process. In particular, we approximate history-dependent policies of team members with state- and belief- (latent state) dependent policy, i.e., $\pi_i \approx \hat{\pi}_i(a_i | s, x_i)$ and $T_{x_i} \approx \hat{T}_{x_i}(x'_i | x_i, a, s')$. This modeling choice while being an approximation enables us to arrive at a suboptimal but tractable solution to the general problem of modeling team behavior. Given this

modeling choice, the setting of P1 reduces to that of BTIL and allowing application of BTIL to learn approximations of $(b_{x_i}, T_{x_i}, \pi_i) \forall i$ given inputs to P1. Application of BTIL also enables sample- and label-efficient learning, a desirable attribute in practice.

Our implementation of BTIL adopts a modified forward and backward message passing algorithm, which is summarized in the Appendix, by observing that message passing for each agent can be decoupled. This modification reduces the time complexity of the algorithm from $O(h|X|^{2n})$ to $O(nh|X|^2)$, where n denotes the size of team, h is the task horizon, and $|X|$ denotes the cardinality of set of mental states. We conclude this section by reiterating that application of BTIL in TIC results from an approximate but, as evidenced in our experiments, useful modeling choice. Learning history-dependent behavior under partial observability remains an exciting open area of future work. As our approach for generating intervention is modular, its future iterations can adopt advances in partially observable imitation learning to better learn the team model and further improve the quality of interventions.

4.3 Inferring Team Mental States

Leveraging the model learned in P1, we develop a Bayesian inference algorithm to estimate the latent states (x) of the team. To develop this algorithm, we observe that P2 corresponds to a non-linear multi-agent filtering problem when there are *no* interventions. Thus, for the case of no interventions, we can formulate the following forward messages to iteratively compute the distribution of mental states at each time step $p(x_i^t | s^{0:t}, a^{0:t-1})$:

$$\begin{aligned} F(t, x_i = j) &\propto p(x_i^t = j, s^{0:t}, a^{0:t-1}) \\ &= \sum_k F(t-1, k) \hat{T}_{x_i}(x_i^t | x_i^{t-1}, a^{t-1}, s^t) \hat{\pi}_i(a_i^{t-1} | s^{t-1}, x_i^{t-1}) \\ F(0, j) &= b_{x_i}(x_i^0 = j | s^0). \end{aligned} \quad (1)$$

However, when a team receives intervention, the mental state dynamics (\hat{T}_{x_i}) will be affected due to the intervention and thus require modifications to the inference algorithm. Recall that the problem setting focuses on intervention that convey a set of recommended mental states (x) to the team. In practice, it is up to each team member's whether to accept the intervention system's recommendation or not; in this work, to simplify our analysis, we assume that each team member accepts the recommendation with probability p_a . Under this assumption, we next update the inference algorithm for the case when an intervention is provided at time t_{int} , i.e., $z_{t_{int}} = 1$. Even in the case of intervention, the algorithm follows an identical structure to Eq. 1. However, the iterative computation of forward messages is updated as follows when $z_{t_{int}} = 1$,

$$\begin{aligned} F''(t, x_i = j) &\propto p(x_i^t = j, s^{t_{int}+1:t}, a^{t_{int}:t-1} | s^{0:t_{int}}, a^{0:t_{int}-1}) \\ &= \sum_k F'(t-1, k) \hat{T}_{x_i}(x_i^t | x_i^{t-1}, a^{t-1}, s^t) \hat{\pi}_i(a_i^{t-1} | s^{t-1}, x_i^{t-1}), \\ F'(t, x_i) &= \begin{cases} F''(t, x_i), & \text{if } t > t_{int} \\ (1-p_a)p(x_i^t | s^{0:t}, a^{0:t-1}) + p_a \mathbb{1}(x_i^t = x_{int,i}), & t = t_{int} \end{cases} \end{aligned} \quad (2)$$

where $x_{int,i}$ is the mental state recommendation generated by solving sub-problem P3. Having computed the distribution of x using

Eqs. 1-2, a point estimate can be obtained by computing the mode of the distribution: $\hat{x}_i = \arg \max_{x_i} p(x_i | s^{0:t}, a^{0:t-1}) \forall i$.

4.4 Detecting Compatible Mental States

Having estimated the team member's latent mental states, before intervening, a natural question to ask is whether the team is already coordinated. To formalize this question, we consider the problem of learning the mental state compatibility function $g(x|s) : X \times S \rightarrow \mathbb{R}$ that takes task context (s) and team members' mental states (x) as input and returns the degree of compatibility. Observe that the compatibility function is context-dependent, as the same shared preferences may be compatible in some context and incompatible in others. Further, we consider *degree* of compatibility as opposed to a binary notion, as some mental state combinations might be better than others for the team's performance on the task.

In general, deriving $g(\cdot)$ is challenging. Multiple configuration of mental states, which in essence translate to preferences over joint policies or shared plans, can represent perfect coordination. Further, the answer to the question depends on the task specifications (shared reward) and team members' behavior. A resource-intensive approach to arrive $g(\cdot)$ is to involve a domain expert who can provide specification of the compatible combinations of mental states. However, this approach is impractical due to the context-dependent and team-dependent nature of compatibility; the number of states are prohibitively large in practice to allow for manual specification.

Hence, in preparation for our solution for P3, we present a simple metric to approximate $g(\cdot)$. Given the task model and Markovian model of team behavior learned in Sec. 4.2, we define $V_\pi(s, x)$ as

$$V_\pi(s, x) = E_{\pi, T_x} \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | s^t = s, x^t = x \right] \quad (3)$$

where R is the task reward. Observe that V corresponds to the expected cumulative discounted reward (or the V value) of the centralized and fully observable version of the Dec-POMDP task model conditioned on the learned team policy $\hat{\pi}$. Several methods exist for computing V values for the fully observable setting, which scale to large problems and are straightforward to implement [33, 43]. The V value depends on the task model, team behavior, current task context and team mental states. Further, it is straightforward to compute and reflects the impact of team mental states (x) on the team's objective (expected discounted cumulative reward). As V_π fulfills all the feature required of a useful compatibility metric, we use it to estimate $g(x|s) \approx V_\pi(s, x)$. Similarly, we can also derive the most compatible combination of team members' mental states (i.e., shared preferences which help achieve the perfect coordination) as $x_{int} = \arg \max_x V_\pi(s, x)$.

4.5 Generating Effective Interventions

With an approach to compute compatibility of team members' mental states, we discuss the final component of TIC: strategies to resolve the cost and benefit trade-off of intervention. While we seek optimal intervention strategies, computationally solving P3 and deriving them is intractable. Hence, we propose a suite of heuristic strategies to compute z_t given the problem inputs. The proposed heuristics can be categorized as deterministic or stochastic, based on their utilization of inference derived in Sec. 4.3.

4.5.1 Deterministic Strategies. The first set of strategies rely on a point estimate of the mental state \hat{x} and are denoted using intervention functions $f(\hat{x}|s) : X \times S \rightarrow \{0, 1\}$, which assume the current task context (s) and estimated mental state (\hat{x}) as inputs. We present two deterministic strategies: rule-based and value-based.

Rule-based. When domain knowledge is available, we can rely on manual specifications of compatible mental states. For this strategy, we assume a domain expert provides a set of compatible mental states denoted by C_s . Given this input, the strategy simply corresponds to verifying if the current estimate of mental state is compatible and intervening if it is not:

$$f_{rule}(\hat{x}|s) = \mathbb{1}(\hat{x} \notin C_s) \quad (4)$$

If the decision is to intervene, the strategy recommends one of the compatible mental states from C_s to the team.

Value-based. In cases where domain knowledge is absent, we rely on the estimate of $g(x|s) \approx V_\pi(x, s)$ to compute z . First, we quantify the benefit of providing an intervention:

$$\text{regret}(x|s) = \max_{x'} g(x'|s) - g(x|s) \quad (5)$$

$$\text{benefit}(x|s) = \text{regret}(x|s) - c. \quad (6)$$

Given an estimate of x , the strategy decides to intervene when benefit of the intervention exceeds a threshold δ :

$$f_\delta(\hat{x}|s) = \mathbb{1}(\text{benefit}(\hat{x}|s) > \delta) \quad (7)$$

δ is a domain-dependent hyperparameter, which can help tune the average number of interventions generated by the strategy. If the decision to intervention is positive, the team is recommended the most compatible mental state: $x_{int} = \arg \max_x g(x|s)$.

4.5.2 Stochastic Strategies. Instead of relying on the point estimate of x , the second set of strategies utilize its inferred distribution $p(x)$. We denote these strategies using intervention function $\Phi(p|s) : \Delta \times S \rightarrow \{0, 1\}$, where Δ denotes the probability simplex. We present two heuristics to convert any deterministic strategy f (e.g., rule- or value-based) to its probabilistic counterpart Φ_f .

Confidence-based. Denote $\hat{x} = \arg \max_x p(x)$. To avoid spurious interventions, the first heuristic, avoids intervening when the confidence in the estimate is lower than a threshold θ :

$$\Phi_{f,\theta}(p|s) = f(\hat{x}|s) \cdot \mathbb{1}(p(\hat{x}) > \theta) \quad (8)$$

Expectation-based. The second strategy computes a soft decision, by computing the expected value of the deterministic strategy:

$$\Phi_{f,E}(p|s) = \mathbb{1}(E_p[f(x|s)] > 0.5) \quad (9)$$

An intervention is made if the soft decision is greater than half.

5 EXPERIMENTS

We evaluate our algorithm in four synthetic scenarios – *Movers*, *Cleanup*, *Flood*, and *Blackout*– each inspired by real-world applications and with large state space [18, 39]. All domains involve partial observability, require collaboration, and include multiple sub-tasks. Team members’ mental state (x) correspond to their preference over which sub-task to complete next. To complete the task efficiently, the team needs to coordinates its next sub-tasks.

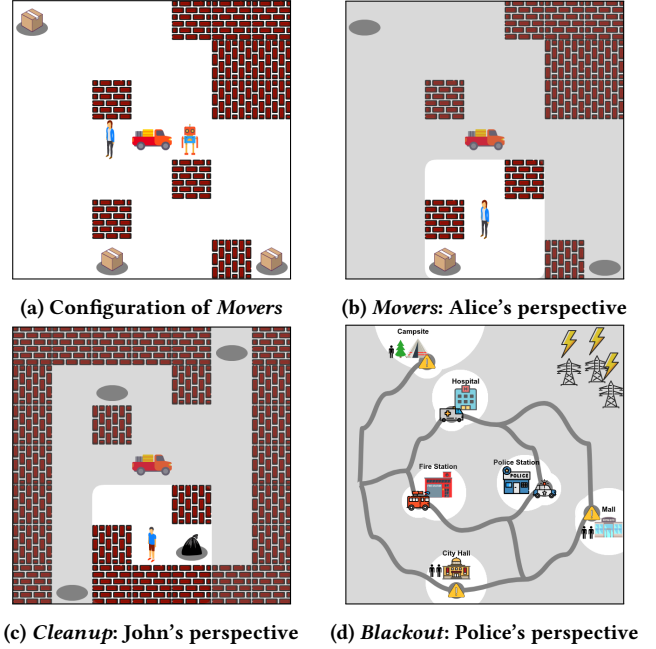


Figure 3: Movers, Cleanup and Blackout domains. Agents can observe only part of the environment (unshaded region).

5.1 Domains

Movers. This domain is similar to Mover’s domain in [39] but adds the challenge of partial observability. As shown in Fig. 3a, both agents, Alice and Rob, start their task with a truck in between and have to move the boxes scattered across the space to the truck. As depicted in Fig. 3b, however, they can only observe their neighboring (unshaded) region. The boxes are heavy, requiring two agents to lift them together. The truck is only accessible while carrying a box, so it is difficult for agents to know where their teammates are at the beginning. An agent can move in one of four cardinal directions, stay put, or attempt to pick up or drop off a box. Each team member can select its next sub-task (x) as one of the three pick-up locations or the drop-off location.

Cleanup. This domain is similar to Cleanup domain in [39] but has a different configuration and enforces partial observability to each agent. A team of two agents, John and Rob, is tasked with moving all the trash bags to the truck. Unlike *Movers*, trash bags can only be picked up by one agent. The set of an agent’s actions is the same as *Movers*. While performing the task, an agent may choose one of five mental states: the pickup and drop-off locations.

Flood. This domain implements the running example in Sec. 1.1. It is an alternate version of Emergency Response domain of Kim and Shah [18] with the added condition of partial observability. The number of people rescued via “rescue” action is different for each disaster site: one at the city hall, two at the campsite, and four in the mall. The goal is to rescue as many people as possible within thirty steps. Each team member can select its next sub-task (x) as one of the four disaster sites.

Blackout. This domain is similar to *Flood* domain but has a different map. In this domain, a team of three members – a police officer, a firefighter, and an EMT(Emergency Medical Technician) – is tasked with rescuing people in the *Blackout* emergency. Similar to *Flood*, each member can only observe their teammates who are at the same location or at landmarks. While the person at the campsite can be rescued by one member, to rescue people in the mall and the city hall, at least two members should work together. Also, the number of people rescued at each site is different: one at the campsite, two in the city hall, and two in the mall. The team is tasked to rescue as many people as possible within 15 steps.

5.2 Data Generation

To evaluate our approach, we created artificial agents and generated a synthetic dataset by simulating teamwork among these agents. For each domain, we implemented a transition of the world T_s and designed each agent’s ground truth policy π_i and their mental state transition T_{x_i} . Each member’s ground truth policy is computed via value iteration by assigning appropriate rewards according to each mental state. The mental state transition is hand-crafted. Due to partial observability, agents make decisions based on their own point estimates of the state \hat{s} , which may differ from actual task states. In situations where teammates are not observable, their actions also have no effect on updating an agent’s mental state. Task execution data are generated by randomly sampling the initial mental state x_i at each member, and iteratively sampling each member’s action $a_i \sim \pi_i(\cdot|\hat{s}, x_i)$, the next state $s' \sim T_s(\cdot|s, a)$, and the next mental state of each member $x'_i \sim T_{x_i}(\cdot|x_i, \hat{s}, s')$ until the completion of the task. For each domain, we generate 500 demonstrations for training BTIL and 100 for evaluating our inference algorithm.

5.3 Evaluation and Metrics

Inference Algorithm. In practice, collecting data of team’s task execution ($s^{0:h}, a^{0:h}$) and annotating team member’s mental states $x^{0:h}$ is resource intensive. To empirically assess the sample- and label-complexity of our inference algorithm, we evaluate its performance using different numbers of training data with different supervision of mental states: 150 demonstrations with 100% supervision, 500 demonstrations with 30% supervision (i.e., 150 labeled demonstrations), and 500 demonstrations with 100% supervision. For each episode, we compute the accuracy of inference as follows: $\frac{1}{nh} \sum_{i=1}^n \sum_{t=0}^h \mathbb{1}(x_i^t = \hat{x}_i^t)$. Inference accuracy evaluates the effectiveness of our forward prediction algorithm, which utilizes the learned behavioral models, to solve P2.

Overall Approach. Due to its modular nature, TIC can utilize any of the intervention strategies proposed in Sec. 4.5. We benchmark multiple versions of TIC, each with different intervention strategy, against two baselines: *no intervention* (i.e., decentralized execution without any intervention) and *centralized policy* (i.e., an optimal team policy which assumes team receives interventions at each step without any consideration of intervention costs). In these experiments, we conduct 100 episodic trials for each combination of domain and intervention strategy and report three metrics: cumulative task rewards, the number of interventions, and problem objective J . TIC computes the decision to intervene z_t using observable data of the team’s task execution ($s^{0:t}, a^{0:t-1}$) and learned

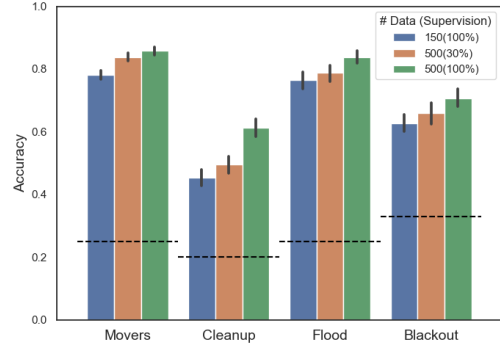


Figure 4: Accuracy of inferring mental states (the dashed black line represents the accuracy of a random guess).

model of team behavior. To show that TIC can work with less supervision, we use models of team behavior learned using semi-supervised data (i.e., 500 data with 30% supervision). If TIC decides to intervene, then the team members update their mental state x with the one recommended by the approach (i.e., $p_a = 1$).

For *Movers* and *Cleanup* domains, where the reward function reflects time to complete the task, we set cost of intervening as $c = 1$. For the emergency response domains, where the reward function corresponds to number of people successfully rescued, defining intervention cost is an ill-defined question; for these domains, we set $c = 0$ and study the trade-off between task reward and number of interventions. We also conduct ablation studies to compare the effect of intervention strategy and hyperparameters. Rule-based strategy is applicable only to those domains where compatibility is easy to define manually. Hence, we apply the rule-based strategy only to *Movers* but evaluate the value-based strategy on all domains. We evaluate the proposed deterministic strategies and their two stochastic versions, by varying the hyperparameters: benefit threshold δ and the certainty threshold θ .

5.4 Results

Accuracy of Task-time Inference of Mental States. As shown in Fig. 4, our inference algorithm predicts team members’ mental states with over 80% accuracy in *Movers* and *Flood*, and around 60% accuracy in *Cleanup* and *Blackout*. In all domains, the algorithm significantly outperforms the probability of random guessing and the performance increases with more data and supervision. We posit the lower accuracy in *Cleanup* and *Blackout* results due to relative lack of training data. *Cleanup* has a much larger state space (~135k) than other domains (*Movers*: ~40k, *Flood*: ~16k, and *Blackout*: ~74k), resulting in less training data relative to the size of the problem. *Blackout* has a much smaller average episode length (≈ 95 for *Movers*, ≈ 15 for *Blackout*), resulting in less training data per demonstration. These results suggest that the proposed approach is capable of learning team behavioral models and inferring mental states with a reasonable accuracy and that, with sufficient training data, the effects of modeling assumptions are not severe.

Utility of Automated Task-time Interventions. The baselines *no intervention* and *centralized policy* can be viewed as two ends of

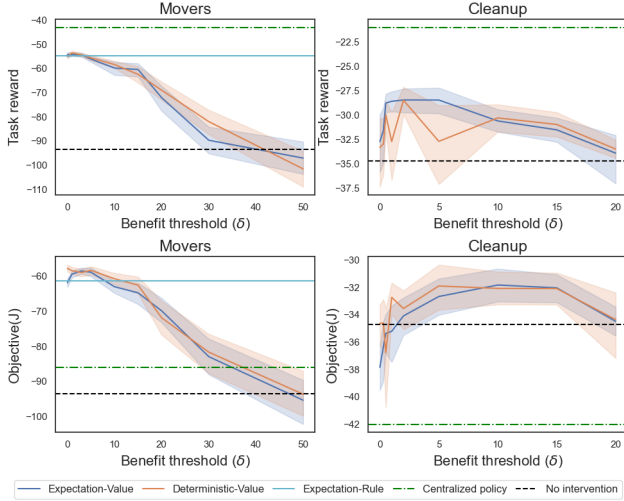


Figure 5: Team’s cumulative task reward (top) and the problem objective J (bottom) in *Movers* and *Cleanup* domains.

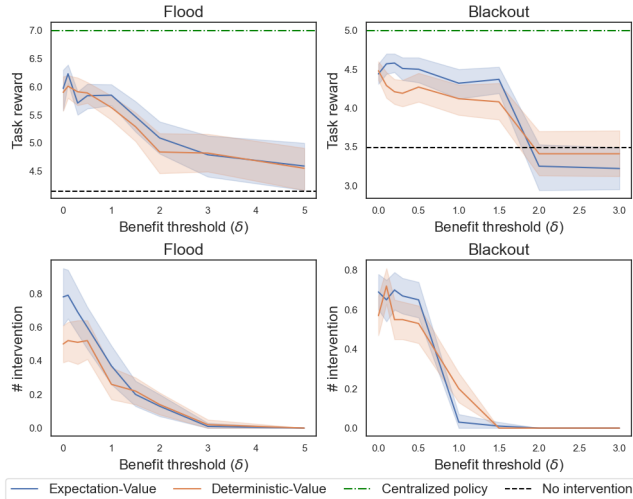


Figure 6: Task reward (top) and the number of interventions (bottom) in *Flood* and *Blackout* domains.

the spectrum of the solution space. Unsurprisingly, the *centralized policy* results in the highest task reward. However, as depicted in Fig. 5 (bottom) and reflected in the problem objective J , this high performance comes at a much higher total cost of interventions. In practice, it is typically infeasible to always intervene either due to environment constraints or intervention overheads. On the other end of the spectrum, teams that receive no intervention do not incur its costs but also perform poorly. In comparison and as depicted in Figs. 5–6, TIC (with a suitable choice of hyperparameter δ) successfully helps improve team’s task performance for all domains with relatively small number of timely but impactful interventions. Encouragingly, TIC is able to generate these interventions despite using noisy inference of team members’ mental states.

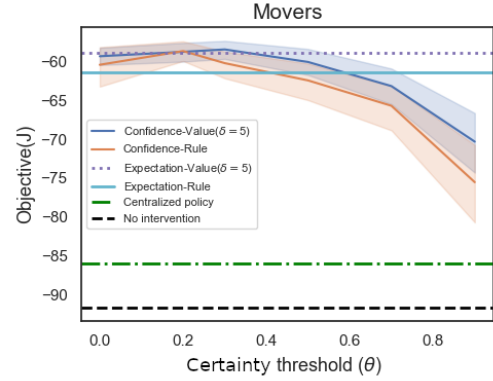


Figure 7: Objective J (with $c = 1$) in the *Movers* domain.

Effect of Estimating Compatibility. We observe in Fig. 7 that the proposed value-based approach to estimating team compatibility $g(\cdot)$ can achieve the same or better performance as the rule-based approach. This suggests that the value V_π defined in Eq. 3 is a good proxy for compatibility $g(x|s)$ and enables an objective task-time assessment of teamwork in domains where hand-crafting the compatibility metric may be resource intensive.

Effect of Intervention Strategy and Hyperparameters. Our ablation experiments comparing different versions of TIC show that even though TIC can improve teamwork, determining the most optimal intervention strategy is difficult due to its dependence on a large number of factors such as task reward, team behavior, inference uncertainties, and intervention costs. Through Figs. 5-7, we also assess the effect of intervention strategy and its hyperparameters (δ, θ) on the overall problem objective (J). We observe that there is no clear winner among proposed intervention strategies and, even with fixed interruption cost, the relationship of objective J with both hyperparameters and number of interventions is non-linear. Fortunately, our computational approach also offers a mechanism for selecting the appropriate intervention strategy through simulation. In practical applications of this and similar approaches, we suggest first collecting data of past execution data of real world team’s behavior, using this data to learn team model and simulate team behavior (similar to our experiment protocol), and then through simulation conduct a hyperparameter analysis to empirically select the most suitable intervention strategy and its hyperparameters.

6 CONCLUSION

We present TIC, an algorithm for generating automated task-time interventions to improve teamwork. Using past and current task execution data, TIC detects poor coordination during teamwork and uses heuristics to trade-off costs and benefits of interventions. We conduct simulation experiments where multi-agent teams perform sequential tasks in partially observable environments and receive interventions generated using TIC. The results show that automated interventions can successfully improve team performance and provide proof-of-concept for development of automated intervention systems to improve human-human and human-AI teamwork.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive feedback. This work was supported in part by NSF award #2205454. Sangwon Seo was partially supported by the Army Research Office through Cooperative Agreement Number W911NF-20-2-0214 and Rice University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

- [1] Stefano V Albrecht and Peter Stone. 2018. Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence* 258 (2018), 66–95.
- [2] Ofra Amir, Barbara J Grosz, and Krzysztof Z Gajos. 2016. MIP-nets: enabling information sharing in loosely-coupled teamwork. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [3] Chris Brady, Karl Tuyls, and Shayegan Omidshafiei. 2022. *AI for Sports*. CRC Press.
- [4] Emily Britton, Natalie Simper, Andrew Leger, and Jenn Stephenson. 2017. Assessing teamwork in undergraduate education: a measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education* 42, 3 (2017), 378–397.
- [5] Sharolyn Converse, JA Cannon-Bowers, and E Salas. 1993. Shared mental models in expert team decision making. *Individual and group decision making: Current issues* 221 (1993), 221–46.
- [6] Roger D Dias, Marco A Zenati, Heather M Conboy, Lori A Clarke, Leon J Osterweil, George S Avrunin, and Steven J Yule. 2021. Dissecting cardiac surgery: A video-based recall protocol to elucidate team cognitive processes in the operating room. *Annals of surgery* (2021).
- [7] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems* 29 (2016).
- [8] Dennis Granäsén. 2019. Towards automated assessment of team performance by mimicking expert observers' ratings. *Cognition, Technology & Work* 21, 2 (2019), 253–274.
- [9] Gregory M Gremillion, Daniel Donavanik, Catherine E Neubauer, Justin D Brody, and Kristin E Schaefer. 2018. Estimating human state from simulated assisted driving with stochastic filtering techniques. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 113–125.
- [10] Barbara J Grosz and Sarit Kraus. 1996. Collaborative plans for complex group action. *Artificial Intelligence* 86, 2 (1996), 269–357.
- [11] Laura M Hiatt, Cody Narber, Esube Bekele, Sangeet S Khemlani, and J Gregory Trafton. 2017. Human modeling for human–robot collaboration. *The International Journal of Robotics Research* 36, 5-7 (2017), 580–596.
- [12] Guy Hoffman. 2019. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218.
- [13] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. 2013. Stochastic variational inference. *Journal of Machine Learning Research* (2013).
- [14] Eric Horvitz and Johnson Apacible. 2003. Learning and reasoning about interruption. In *Proceedings of the 5th international conference on Multimodal interfaces*. 20–27.
- [15] Jiechuan Jiang and Zongqing Lu. 2018. Learning attentional communication for multi-agent cooperation. *Advances in neural information processing systems* 31 (2018).
- [16] Mingxuan Jing, Wenbing Huang, Fuchun Sun, Xiaojian Ma, Tao Kong, Chuang Gan, and Lei Li. 2021. Adversarial option-aware hierarchical imitation learning. In *International Conference on Machine Learning*. PMLR, 5097–5106.
- [17] Ece Kamar, Ya'akov Gal, and Barbara J Grosz. 2009. Incorporating helpful behavior into collaborative planning. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. Springer Verlag.
- [18] Joseph Kim and Julie A Shah. 2016. Improving team's consistency of understanding in meetings. *IEEE Transactions on Human-Machine Systems* 46, 5 (2016), 625–637.
- [19] Ben D Lawson, Thomas W Britt, Amanda M Kelley, Jeremy R Athy, and Shauna M Legan. 2017. Computerized tests of team performance and crew coordination suitable for military/aviation settings. *Aerospace Medicine and Human Performance* 88, 8 (2017), 722–729.
- [20] Hoang M Le, Yisong Yue, Peter Carr, and Patrick Lucey. 2017. Coordinated multi-agent imitation learning. In *International Conference on Machine Learning*. PMLR, 1995–2003.
- [21] Linda Searle Leach, Robert C Myrtle, Fred A Weaver, and Sriram Dasu. 2009. Assessing the performance of surgical teams. *Health care management review* 34, 1 (2009), 29–41.
- [22] Namhoon Lee and Kris M Kitani. 2016. Predicting wide receiver trajectories in american football. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–9.
- [23] Sea Ling and Seng Wai Loke. 2003. MIP-Nets: A compositional model of multi-agent interaction. In *International Central and Eastern European Conference on Multi-Agent Systems*. Springer, 61–72.
- [24] Liansie Mingyue Ma, Martijn IJtsma, Karen M Feigh, and Amy R Pritchett. 2022. Metrics for Human-Robot Team Design: A Teamwork Perspective on Evaluation of Human-Robot Teams. *ACM Transactions on Human-Robot Interaction* (2022).
- [25] Patrick MacAlpine and Peter Stone. 2017. Evaluating ad hoc teamwork performance in drop-in player challenges. In *International Conference on Autonomous Agents and Multiagent Systems*. Springer, 168–186.
- [26] Jean MacMillan, Elliot E Entin, and Daniel Serfaty. 2004. Communication overhead: The hidden cost of team cognition. (2004).
- [27] Daniele Masato, Timothy J Norman, Wamberto W Vasconcelos, and Katia Sycara. 2011. Agent-oriented incremental team and activity recognition. In *Twenty-Second International Joint Conference on Artificial Intelligence*.
- [28] Reuth Mirsky, Ignacio Carlucho, Arrasy Rahman, Elliot Fosong, William Macke, Mohan Sridharan, Peter Stone, and Stefano V Albrecht. 2022. A Survey of Ad Hoc Teamwork Research. In *European Conference on Multi-Agent Systems (EUMAS)*.
- [29] Julia Neily, Peter D Mills, Yinong Young-Xu, Brian T Carney, Priscilla West, David H Berger, Lisa M Mazzia, Douglas E Paull, and James P Bagian. 2010. Association between implementation of a medical team training program and surgical mortality. *JAMA* 304, 15 (2010), 1693–1700.
- [30] Adam Norton, Henny Admoni, Jacob Crandall, Tesca Fitzgerald, Alvika Gautam, Michael Goodrich, Amy Saretsky, Matthias Scheutz, Reid Simmons, Aaron Steinfeld, et al. 2022. Metrics for Robot Proficiency Self-assessment and Communication of Proficiency in Human-robot Teams. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–38.
- [31] Frans A Oliehoek and Christopher Amato. 2016. *A concise introduction to decentralized POMDPs*. Springer.
- [32] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, Jan Peters, et al. 2018. An algorithmic perspective on imitation learning. *Foundations and Trends® in Robotics* 7, 1-2 (2018), 1–179.
- [33] Martin L Puterman. 2014. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.
- [34] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. 2016. Planning for autonomous cars that leverage effects on human actions.. In *Robotics: Science and systems*, Vol. 2. Ann Arbor, MI, USA, 1–9.
- [35] Eduardo Salas, Denise L Reyes, and Susan H McDaniel. 2018. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist* 73, 4 (2018), 593.
- [36] Eduardo Salas, Scott Tannenbaum, Deborah Cohen, and Gary Latham. 2013. *Developing and enhancing teamwork in organizations: Evidence-based best practices and guidelines*. Vol. 33. John Wiley & Sons.
- [37] Suchi Saria and Sridhar Mahadevan. 2004. Probabilistic Plan Recognition in Multiagent Systems.. In *ICAPS*. 287–296.
- [38] Sangwon Seo, Lauren R Kennedy-Metz, Marco A Zenati, Julie A Shah, Roger D Dias, and Vaibhav V Unhelkar. 2021. Towards an AI coach to infer team mental model alignment in healthcare. In *2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE, 39–44.
- [39] Sangwon Seo and Vaibhav V. Unhelkar. 2022. Semi-Supervised Imitation Learning of Team Policies from Suboptimal Demonstrations. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, Lud De Raedt (Ed.). International Joint Conferences on Artificial Intelligence Organization, 2492–2500. <https://doi.org/10.24963/ijcai.2022/346> Main Track.
- [40] Arjun Sharma, Mohit Sharma, Nicholas Rhinehart, and Kris M Kitani. 2018. Directed-info gail: Learning hierarchical policies from unsegmented demonstrations using directed information. *arXiv preprint arXiv:1810.01266* (2018).
- [41] Herbert A Simon. 1990. Bounded rationality. In *Utility and probability*. Springer, 15–18.
- [42] Jiaming Song, Hongyu Ren, Dorsa Sadigh, and Stefano Ermon. 2018. Multi-agent generative adversarial imitation learning. In *Advances in Neural Information Processing Systems*. 7472–7483.
- [43] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- [44] Milind Tambe, Emma Bowring, Hyuckchul Jung, Gal Kaminka, Rajiv Maheswaran, Janusz Marecki, Pragnesh Jay Modi, Ranjit Nair, Stephen Okamoto, Jonathan P Pearce, et al. 2005. Conflicts in teamwork: Hybrids to the rescue. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*. 3–10.
- [45] Yla R Tausczik and James W Pennebaker. 2013. Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 459–468.

- [46] Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes, Ian Graham, William Spearman, Tim Waskett, Dafydd Steel, et al. 2021. Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research* 71 (2021), 41–88.
- [47] Vaibhav V Unhelkar and Julie A Shah. 2019. Learning models of sequential decision-making with partial specification of agent behavior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 2522–2530.
- [48] Piet Van den Bossche, Wim Gijssels, Mien Segers, Geert Woltjer, and Paul Kirschner. 2011. Team learning: building shared mental models. *Instructional Science* 39, 3 (2011), 283–301.
- [49] Chen Wang, Claudia Pérez-D’Arpino, Danfei Xu, Li Fei-Fei, Karen Liu, and Silvio Savarese. 2021. Co-GAIL: Learning Diverse Strategies for Human-Robot Collaboration. In *5th Annual Conference on Robot Learning*.
- [50] Rena Xu, Matthew J Carty, Dennis P Orgill, Stuart R Lipsitz, and Antoine Duclos. 2013. The teaming curve: a longitudinal study of the influence of surgical team familiarity on operative time. *Annals of surgery* 258, 6 (2013), 953–957.
- [51] Shi-guang Yue, Peng Jiao, Ya-bing Zha, and Quan-jun Yin. 2015. A logical hierarchical hidden semi-Markov model for team intention recognition. *Discrete Dynamics in Nature and Society* 2015 (2015).