# AI-Assisted Human Teamwork

**Sangwon Seo**
Rice University
Houston, TX 77005, USA
sangwon.seo@rice.edu

## Abstract

Effective teamwork translates to fewer preventable errors and higher task performance in collaborative tasks. However, in time-critical tasks, successful teamwork becomes highly challenging to attain. In such settings, often, team members have partial observability of their surroundings, incur high cost of communication, and have trouble estimating the state and intent of their teammates. To assist a team in improving teamwork at task time, my doctoral research proposes an automated task-time team intervention system. Grounded in the notion of shared mental models, the system first detects whether the team is on the same page or not. It then generates effective interventions to improve teamwork. Additionally, by leveraging past demonstrations to learn a model of team behavior, this system minimizes the need for domain experts to specify teamwork models and rules.

## Introduction

Effective teamwork is critical across diverse domains, including manufacturing, disaster response, and healthcare. In these domains, tasks are primarily collaborative, and the absence of teamwork often leads to catastrophic accidents (Seo et al. 2021). At the same time, perfecting teamwork is highly challenging. Many times, team members have limited observability over the task environment and possess different thoughts, awareness, or preferences regarding their roles and plans for the task. This problem is particularly pronounced in time-critical tasks where information exchange is not readily available in a team. In certain domains, such as sports, teams often employ human coaches who can holistically monitor ongoing tasks and coordinate team members during task execution. However, in other domains, resource constraints make it difficult to find experts for such roles.

To address the challenge of enhancing teamwork during task execution, my doctoral research aims to develop an automated task-time intervention system for teamwork. Analogous to human coaches, this system monitors the team's behavior in performing a task via sensors and (when it detects imperfect coordination) provides interventions at task time to improve teamwork. Moreover, instead of relying on rule-based approaches, my thesis seeks to utilize past demonstrations of teamwork to build the system. Since demonstrations can be continuously collected by sensors without requiring significant attention from experts, this approach alleviates the tiring and resource-intensive efforts of domain experts needed to hand-craft the features of effective teamwork.

However, constructing such a system is nontrivial. Teamwork is influenced by various factors, and defining what constitutes effective teamwork can vary depending on the context and perspective. In order to simplify the problem and arrive at a practical solution, my thesis focuses on two axes of teamwork: shared mental models and team performance (Bisbey, Traylor, and Salas 2021). Shared mental models are a critical construct of teamwork that pertains to whether a team is on the same page or not. Meanwhile, team performance is defined as the team's achievement regarding shared task objectives and serves as an objective measure of teamwork. As introduced in (Seo et al. 2021), real-world scenarios often witness task failures stemming from a lack of team alignment, underscoring the importance of this approach.

To enable a computational understanding of teamwork, a generative model of teamwork is essential. Specifically, in accommodating humans' intentional behavior, this teamwork model should consider not only behaviors revealed through observation but also each member's unobservable intents. Consequently, learning a teamwork model from intent-driven demonstrations necessitates a specialized imitation learning algorithm. Leveraging this teamwork model, the proposed intervention system assesses team alignment and prescribes interventions as appropriate to improve team performance. To this end, my thesis addresses three interrelated research problems: 1. imitation learning of intent-driven team behavior; 2. detection of insufficient team alignment; and 3. intervention strategies to improve teamwork.

## Imitation Learning of Team Behavior

To capture the influence of intents on decision-making, in Seo and Unhelkar (2022), I explicitly represent the $i$-th agent's policy as $\pi_i(a_i|s, x_i)$ where $a_i$ is an action, $s$ is a task state, and $x_i$ denotes an intent. Additionally, since the agent may update their intent during the task, I model its temporal dynamics as $T_{x_i}(x_i'|s', a, x_i)$ where $x_i', s'$ are the next intent and the next task state, respectively, and $a$ is the joint actions of the team. Most multi-agent imitation learning algorithms cannot be applied to my teamwork model as they consider full observability and only learn $\pi(a|s)$. Hence, I propose

Bayesian Team Imitation Learner (BTIL) which can learn both the intent-aware team policies $\pi(a|s,x)$ and intent dynamics $T_x(x'|s,a,x)$ from demonstrations in collaborative settings (Seo and Unhelkar 2022). Built upon mean-field variational inference, BTIL achieves the sample- and label-efficient learning, which is a notable advantage given that intents are not observable.

I am currently working on an extension of BTIL to handle more complex domains. Since BTIL is based on a Bayesian approach, it struggles to scale up to domains with high-dimensional or continuous state spaces. On the other hand, the extension represents a teamwork model using function approximators and learns the accurate model by augmenting it with online samples. In a single-agent scenario, Option-GAIL (Jing et al. 2021) can be a candidate to train the agent model. However, Option-GAIL relies on adversarial training, which often results in unstable optimization. Thus, my plan is to develop a non-adversarial-training imitation learning algorithm for a single agent and then extend it to the multi-agent collaborative setting.

## Detection of Imperfect Coordination

My research proposes methods to assess teamwork based on the notion of shared mental models, which enables implicit coordination in a team (Bisbey, Traylor, and Salas 2021). When a team operates with a shared mental model, each member continuously adapts their intent during task execution, ensuring compatibility with the intents of others. Thus, my research defines imperfect coordination as a situation where intents between team members are incompatible and proposes methods to measure the compatibility of intents. In Seo et al. (2021), I introduce teamwork scenarios where the misalignment of intents in a team can result in fatal accidents. In Seo, Han, and Unhelkar (2023), I consider general teamwork scenarios where intents can dynamically change during the task. Leveraging task objectives and the teamwork model, I present a method to measure the compatibility of intents without involving domain experts.

To enable this shared intent-based teamwork assessment, it is critical to correctly infer intents of each team member. However, even in presence of teamwork models learned using multi-agent imitation learning (such as BTIL), novel methods are needed to infer team intents in presence of task-time interventions. In training demonstrations used for imitation learning, intents change based on the intent transition $T_x$ alone, whereas during the task execution phase, team members' intents will also be influenced by the interventions. To reflect this effect, my research presents an effective algorithm to infer team members' intents that explicitly considers the interaction between the team and intervention system (Seo, Han, and Unhelkar 2023).

## Intervention Strategies to Improve Teamwork

Based on the assessment on team coordination, my research proposes TIC, an automated Task-time Intervention for improving Collaboration (Seo, Han, and Unhelkar 2023). Interventions can enhance teamwork by rectifying misalignment in team members' intents. However, too many interventions may hinder a team's task execution and cause unintended side effects. Thus, I formulate the objective of TIC to effectively balance the costs and benefits of interventions. Assuming the set of intents are known *a priori*, when decided to intervene, the intervention system directly informs the team of the most compatible intents at the current state. Since solving the TIC objective is computationally intractable, my research presents a set of heuristic strategies that consider both the distribution over the inferred intents and their compatibility.

An immediate future study is to apply TIC to human teams in practice. Under Institutional Review Board (IRB) approval, I have already collected demonstrations from human participants for training TIC. A separate experimental protocol related to validating TIC was recently approved, and the experiment interface is mostly developed. Another future research is to provide effective interventions without relying on prior knowledge of intents. Oftentimes, it is unrealistic to assume the set of possible intents is accessible. In this case, the intervention system cannot directly convey the team its optimal intents. Instead, the system needs to provide interventions using user-interpretable advice, such as the best actions at that moment (Gan et al. 2022). I plan to utilize Bayesian methods to model the update of intents according to these indirect interventions and seek to learn near-optimal intervention strategies.

## Acknowledgments

I thank my advisor, Vaibhav Unhelkar, for his invaluable advice and support. This research was supported in part by NSF award #2205454, the Army Research Office through Cooperative Agreement Number W911NF-20-2-0214, and Rice University funds.

## References

Bisbey, T.; Traylor, A.; and Salas, E. 2021. Transforming teams of experts into expert teams: eight principles of expert team performance. *Journal of expertise*, 4(2).

Gan, J.; Majumdar, R.; Radanovic, G.; and Singla, A. 2022. Bayesian persuasion in sequential decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 5025–5033.

Jing, M.; Huang, W.; Sun, F.; Ma, X.; Kong, T.; Gan, C.; and Li, L. 2021. Adversarial option-aware hierarchical imitation learning. In *International Conference on Machine Learning*, 5097–5106. PMLR.

Seo, S.; Han, B.; and Unhelkar, V. 2023. Automated Task-Time Interventions to Improve Teamwork using Imitation Learning. *arXiv preprint arXiv:2303.00413*.

Seo, S.; Kennedy-Metz, L. R.; Zenati, M. A.; Shah, J. A.; Dias, R. D.; and Unhelkar, V. V. 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)*, 39–44. IEEE.

Seo, S.; and Unhelkar, V. V. 2022. Semi-Supervised Imitation Learning of Team Policies from Suboptimal Demonstrations. *arXiv preprint arXiv:2205.02959*.