RL-HAT: A New Framework for Understanding Human-Agent Teaming

Kiana Jafari Meimandi [1], **Matthew Bolton** [1], **Peter Beling** [2]

[1] Department of Systems and Information Engineering, University of Virginia
[2] National Security Institute, Virginia Tech
kj6vd@virginia.edu, mlb4b@virginia.edu, beling@vt.edu

Abstract

This paper presents a novel framework for human-agent teaming grounded in the principles of Reinforcement Learning (RL). Recognizing the need for a unified language across various disciplines, we utilize RL concepts to provide a standard for the understanding and evaluation of diverse teaming strategies. Our framework extends beyond traditional RL constructs, integrating aspects such as belief states, prior knowledge, social considerations, situational awareness, and mental models. A particular focus is placed on the role of ethics and trust in effective teaming. Additionally, we discuss how sensor data, perception models, and actuator modules can be incorporated, emphasizing the adaptability of our framework to a broad range of tasks and environments. We believe this work forms a substantial contribution to the field of human-agent teaming, establishing a solid foundation for future research and application.

Introduction

Over the past few decades, human-agent teaming (HAT) has undergone significant development, drawing contributions from multiple disciplines, including psychology (Lyons et al. 2021), human factors engineering (Johnson et al. 2021), cognitive sciences (Khamassi et al. 2018), neuroscience (Lakhmani et al. 2016), computer science, and robotics (Khavas, Ahmadzadeh, and Robinette 2020; Hani Daniel Zakaria et al. 2021). Despite the interconnections between these fields, a lack of a unified perspective persists. Additionally, while artificial intelligence (AI) has made notable advancements, human operators working alongside AI frequently encounter partners who excel in certain areas but lack fundamental skills in others (Dellermann et al. 2019). Consequently, it is essential to develop a comprehensive HAT framework for engineering AI capable of managing and comprehending the complementary capabilities of human operators that is helpful for both researchers and practitioners.

Several HAT frameworks stemming from various disciplines have been proposed in the literature. These highlight the importance of effective communication, collaboration, and coordination. These frameworks have been applied in diverse domains such as military operations (Nothwang et al. 2016), search and rescue missions (Tai 2021), and industrial automation (Schelble, Flathmann, and McNeese 2020). The HACO (Human-AI Collaboration) framework adopts a software development perspective and advocates for a model-driven approach to HAT system development via a graphical user interface (Dubey et al. 2020). Several frameworks in the literature originate from human factors engineering and focus on design (Cooke, Demir, and Huang 2020), human cognition (Scheutz, DeLoach, and Adams 2017), team behavior (Aldridge and Bethel 2023; Ma et al. 2022), and composition (Lohani et al. 2017; Schelble et al. 2020). Moreover, there are frameworks that specifically address certain applications and domains. For instance, BO-MUSE employs Bayesian methods to optimize the experimental design involving human-AI teams (Gupta et al. 2023). Evertsz and Thangarajah employ TDF-T diagrams to engineer human-agent teams and provide context-specific team cognition to human team members during the runtime of the StarCraft strategy game (Evertsz and Thangarajah 2020).

The selection of an appropriate framework heavily relies on the specific context and objectives of the HAT scenario, and the efficacy of a framework hinges on its alignment with the team's requirements. To the best of our knowledge, no single framework has demonstrated adaptability to different settings. Furthermore, there is a lack of literature on evaluation methods for HAT frameworks. This paper introduces a novel HAT framework that is based on reinforcement learning (RL) with the aim of establishing a shared language and conceptual framework that can be employed across disciplines. The ultimate objective is to foster improved collaboration and communication among researchers and practitioners from diverse fields. This should lead to fresh insights and innovative advancements in the field of HAT.

Reinforcement Learning

Reinforcement learning (RL) involves learning how to make decisions or mapping situations to actions while interacting with the environment in a way that maximizes a numerical reward signal. Trial-and-error search and delayed reward are the key defining features of reinforcement learning (Sutton, Barto et al. 1998). In a decision-making task, the signal indicating the agent's choices is action ($a_t \in A$), the one signifying the context in which those choices are made is observation ($o_t \in O$), and the one that defines the agent's ultimate

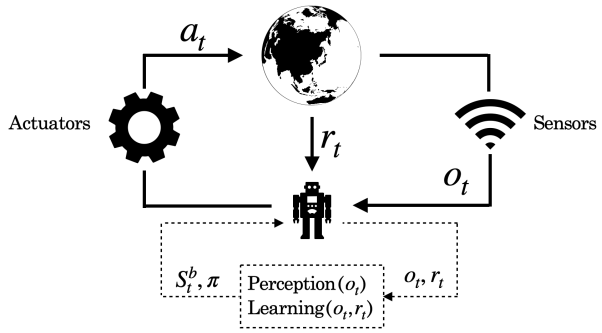

Figure 1: Single-agent reinforcement learning

objective is the reward (r_t). Anything the agent cannot arbitrarily change is considered to be part of its environment. The agent-environment boundary indicates the limit of the agent's control rather than its knowledge and can vary depending on the intended purpose. In a complicated environment, many agents may operate simultaneously, each within its boundary (Sutton, Barto et al. 1998).

Three main sub-components of RL are *policy*, *value function*, and a *model* of the environment. A policy, $\pi : O \rightarrow \Delta(A)$, is the agent's behavior or a mapping from its perception of the state (s_t^b) of the environment to actions. The value function, V, predicts future rewards or how good it is to be in the current state (and taking action A) in the long run. A model of the environment simulates its behavior and predicts the consequences given the state and action(s) taken by the agent.

Multi-agent reinforcement learning (MARL) extends sequential decision-making to multiple agents interacting within a shared environment and potentially with each other (Gronauer and Diepold 2022). Joint actions of all agents impact environment state changes and individual reward signals. Agents optimize their own long-term rewards based on other agents' policies. MARL can be centralized, where a central agent makes decisions for all agents, allowing better coordination but with increased complexity and potential single point of failure. Alternatively, it can be decentralized, with each agent having its own policy, leading to scalability and robustness (Zhang, Yang, and Başar 2021a; Sharma et al. 2021). Emergent behavior occurs due to interactions and individual learning algorithms, allowing decentralized coordination and cooperation without a central controller (Martinez-Gil, Lozano, and Fernández 2017). It is desirable for efficiency, flexibility, and adaptability in MARL, though designing and controlling it can be challenging. Examples of emergent behavior include communication protocols, specialized roles, and the self-organization of collective behaviors (Gupta, Hazra, and Dukkipati 2020).

HAT and RL

Human-agent teaming refers to the collaborative effort between one or more humans and autonomous agents to achieve common goals. It is characterized by interdependence in activity and outcomes, where each team member, whether human or autonomous agent, has a unique role and is recognized as such (O'Neill et al. 2022). The agent in HAT always has some level of autonomy, meaning it can take action and make decisions within the context of the overall team and its goals (O'Neill et al. 2022).

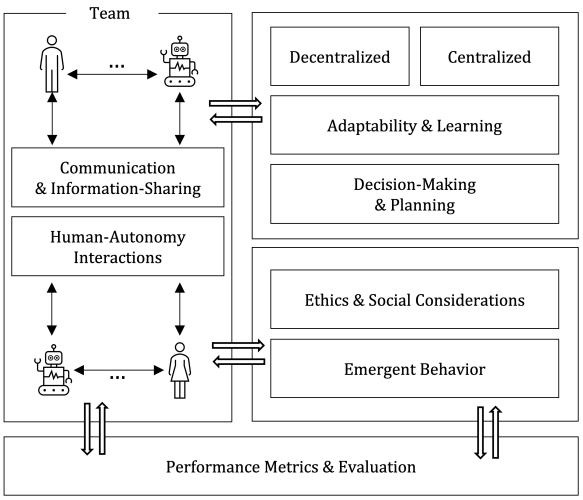

Figure 2: HAT components in the context of teaming.

Human-agent teaming can be categorized into six main components: decision-making and planning, communication and information-sharing, human-autonomy interactions, performance, adaptability and learning, and ethics and social considerations (Figure 2). This work argues that all these components can be explained using RL concepts.

Decision-making and Planning: This involves choosing appropriate actions to maximize some notion of cumulative reward. In the RL framework, this is analogous to the policy of an agent, which is a mapping from states to actions. For successful Human-agent teaming, both the human and the agent need to have compatible or cooperative decision-making and planning mechanisms. This could involve considering the reward functions, utilities, or satisficing conditions of both the human and the agent, and ensuring that the combined policy respects these.

Communication and Information-Sharing: In RL, agents need to observe the state of the environment in order to make decisions. In a human-agent team, it's critical that both parties share relevant information about their observations, intentions, and actions. This could be implemented in a number of ways, such as through a shared state representation, through auxiliary communication channels (like textual or spoken language), or even through modifications to the action space to include communication actions.

Human-Autonomy Interactions: This could be understood as a multi-agent extension of RL, where the human and the RL agent are both actors in the environment. The interaction dynamics need to be carefully managed so that both the human and the agent can effectively collaborate. One common approach is to use models that account for the behavior of the other actor, for example, by predicting the human's actions, intentions or preferences.

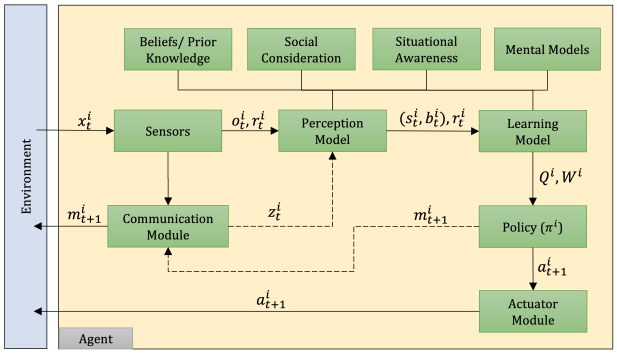

Figure 3: Agent model and its interaction with the environment.

Performance: Just as in RL, performance in a human-agent team is typically measured by some kind of cumulative reward. However, this might be more complex because it could involve multiple, potentially conflicting, objectives. Furthermore, since humans are part of the team, subjective measures like user satisfaction, trust, perceived usefulness, and perceived performance of the RL agent could also be important.

Adaptability and Learning: RL agents are, by definition, capable of learning and adapting over time. In a human-agent team, it's crucial that the agent is able to adapt to the human's behaviors and preferences, and, ideally, the human should adapt to the agent's capabilities. This could involve concepts from transfer of learning (to rapidly adapt to new humans or tasks), inverse reinforcement learning (to learn human preferences), or meta-learning (to learn how to learn from humans).

Ethics and Social Considerations: While this isn't typically a standard focus of RL research, it is critical in any application involving humans. This could involve ensuring that the RL agent respects human values and social norms, and is transparent, interpretable, and fair. Ethical considerations could also shape the reward function or policy of the RL agent, constraining it to make "ethical" decisions or guiding it to promote beneficial outcomes for the human.

This research uses these six components to propose a conceptual framework for designing, implementing, and evaluating the teamwork between humans and artificial agents using reinforcement learning.

Proposed Framework

Using RL terminology, it can be said that in any given scenario, numerous agents can operate simultaneously while having their own objectives and policy for decision-making as they engage with their environment. Their abilities may vary in how they perceive and gather information from the environment, evaluate feedback, and learn new patterns. Agents, as members of a team, have two kinds of boundaries when interacting with the environment; one is the individual boundary, and the other is the team's boundary. Within the team's boundary, agents communicate, share information and interact toward a common goal. Therefore, their personal and team goals influence agents' decision-making processes. Furthermore, agents can access information that would otherwise be unavailable outside of the team, affecting their decision-making process. Figure 3 illustrates a single agent model, its components, and its interaction with the environment.

Beliefs (b_t) are the agent's internal representations of the state of the world or aspects of it, whether accurate or not. An agent forms beliefs based on its perception, learning, and prior knowledge (Poole and Mackworth 2010). In RL, this component is typically formalized as a Partially Observable Markov Decision Process (POMDP) where "belief" is a probability distribution over the set of possible states, representing the agent's uncertainty about the true state of the environment (Spaan 2012). The belief is updated over time based on the agent's observations and actions, often using Bayesian updating. The belief update rule in POMDP is usually given by Bayes' rule, which allows the agent to update its belief based on its latest action and observation. In a fully observable environment (Markov Decision Process or MDP), the agent's belief about the current state is always a distribution with all probability mass on the actual current state, because the agent has complete information in this case.

Prior knowledge represents the agent's pre-existing knowledge or information before it interacts with the environment. It could include things the agent was designed to know or understand, or information it has learned in previous interactions (Poole and Mackworth 2010). In RL, the concept of prior knowledge is often incorporated into the model of the environment or the learning process in a number of ways (e.g. value initialization, inductive bias (Zambaldi et al. 2019)).

Social consideration is the agent's capacity to include social contexts and norms when making decisions or interacting with humans. It includes understanding of things like trust, ethics, social roles, cultural norms, and more (Jackson and Williams 2021). While traditional Reinforcement Learning (RL) algorithms primarily focus on maximizing a reward signal, there is growing interest in extending these algorithms to handle complex social considerations like trust and ethics. For example, using inverse reinforcement learning (IRL) the AI agent aims to infer the human's reward function (Adams, Cody, and Beling 2022) or determine what they find satisficing (Richardson 2017). This could be a way to incorporate human values, trust, ethics, and other social considerations into the agent's behavior.

Situational awareness (SA) is the agent's understanding of the current situation or environment in which it is operating(Feng, Teng, and Tan 2009) and predicts how it will change in the immediate future (Endsley 1995). It includes knowledge of both the physical environment and the social context, including the goals, intentions, and actions of other agents. In RL literature, SA spans across several components of an RL system such as state representation, reward function, policy, value decomposition (Sunehag et al. 2017) and more.

Mental models are the agent's internal models or representations of how the world works, including other agents.

Mental models guide the agent's expectations and predictions, and shape how it interprets sensory input and makes decisions (Denzau, North et al. 1994). In RL, a model of the environment is an understanding or representation of how the environment works. Specifically, the model can predict what the next state and reward will be given the current state and action. This concept of an environmental model in RL aligns well with the notion of a mental model.

Sensors are the hardware or software components that collect data from the agent's environment. For an AI agent, sensors could include things like cameras, microphones, or data input streams. Typically, an RL agent interacts with its environment by taking actions based on its current state which is a representation of what the agent senses or observes about the environment. Depending on the specifics of the RL problem, this state might include a variety of different types of sensory information, such as visual input, audio input, or other types of sensory data.

Perception model is the method or algorithm that the agent uses to interpret the data it collects from its sensors before it is incorporated into its beliefs and mental models (Cangelosi 2010). While the term perception model isn't explicitly used in RL literature, concepts like observations, state representations, state transition functions, observation functions, and neural network-based functions approximators all contribute to the agent's perception of its environment.

Learning model represents the method or algorithm the agent uses to update its beliefs, mental models, and behaviors based on experience. It can include methods like reinforcement learning, supervised learning, unsupervised learning, and more. In RL, policy (π), value function (V or Q), reward function (R), learning algorithms (e.g. Q-learning, SARSA, Policy Gradient methods, Actor-Critic methods) and model of environment all contributes to the learning models (Sutton, Barto et al. 1998).

Communication module is the component of the agent that allows it to communicate with humans or other agents. It might involve natural language processing for a chatbot or a data transmission protocol for a more specialized agent (Balaji and Srinivasan 2010). In a MARL context, communication might be represented as a function or module that transforms an agent's internal state or observations into a message, and/or one that transforms received messages into inputs for the agent's decision-making process. For example, in some cases, an agent might have a communication policy, which is a function that maps its internal state to a message. Similarly, the agent might have a message-processing function that maps received messages to inputs for the decision-making process. In Multi-Agent Reinforcement Learning (MARL), various methods have been proposed for enabling communication between agents. The primary goal of these methods is to allow agents to share information to improve coordination and collective performance (Zhang, Yang, and Başar 2021b).

Policy: In the context of an agent, a policy refers to the strategy or rule that the agent follows when deciding what actions to take in a given situation. The policy is often based on the agent's beliefs, mental models, and learning. This concept aligns with the definition of policy in RL.

Actuator module: This is the component of the agent that carries out actions in the world. For a robot, this could include motors that move its limbs. For a software agent, this could include sending messages or changing data. In RL, the action space (A) can be considered a representation of the capabilities of the actuator module (Sutton, Barto et al. 1998). The action space includes all the actions that an agent can potentially take. For instance, in a robotic arm task, the action space might consist of all possible joint angles or motor torques.

Conclusion and Future Work

In conclusion, our proposed framework offers a novel, robust, and comprehensive approach to engineering autonomous agents in human-agent teaming based on principles of Reinforcement Learning (RL). By harnessing the expressive power of RL, we are able to create a common language that unifies disparate disciplines and provides a consistent, quantifiable means of evaluating and comparing different teaming strategies.

Our framework expands upon traditional RL components, incorporating elements such as belief states, prior knowledge, social consideration, situational awareness, mental models, and communication. These enrichments enable a deeper understanding and modeling of human-agent interaction (largely framed around human-centered concepts), moving beyond simple reward optimization towards more nuanced, context-sensitive behavior.

Importantly, our framework accommodates not only fully observable environments but also partially observable scenarios, with a particular emphasis on handling uncertainties. We've also addressed the critical aspect of communication in multi-agent scenarios, highlighting methodologies such as centralized/decentralized learning, and transfer learning techniques. Furthermore, our approach pays specific attention to ethical considerations and trust, integral factors in real-world applications of human-agent teaming.

The adaptability of our model enables it to incorporate sensor data, perception models, and actuation modules, demonstrating its applicability to a wide range of tasks and environments. At the same time, our framework is grounded in the fundamental elements of RL, namely states, actions, policies, value functions, and reward functions. The combination of traditional RL constructs with advanced social, psychological, and environmental considerations paves the way for more effective and efficient human-agent collaboration.

Future work will explore empirical evaluations of our framework, with a particular focus on real-world applications and scenarios where human-agent teaming is key. We believe that this line of research will contribute significantly to the field, fostering increased understanding and refinement of human-agent teaming methodologies and applications.

We hope that this work will prompt further exploration and innovation in the field of human-agent teaming, moving toward more effective, efficient, and socially-aware collaborative systems.

References

Adams, S.; Cody, T.; and Beling, P. A. 2022. A survey of inverse reinforcement learning. *Artificial Intelligence Review*, 55(6): 4307–4346.

Aldridge, A. L.; and Bethel, C. L. 2023. M-OAT Shared Meta-Model Framework for Effective Collaborative Human-Autonomy Teaming. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 663–666.

Balaji, P. G.; and Srinivasan, D. 2010. An introduction to multi-agent systems. *Innovations in multi-agent systems and applications-1*, 1–27.

Cangelosi, A. 2010. Grounding language in action and perception: From cognitive agents to humanoid robots. *Physics of life reviews*, 7(2): 139–151.

Cooke, N.; Demir, M.; and Huang, L. 2020. A framework for human-autonomy team research. In *Engineering Psychology and Cognitive Ergonomics. Cognition and Design: 17th International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*, 134–146. Springer.

Dellermann, D.; Ebel, P.; Söllner, M.; and Leimeister, J. M. 2019. Hybrid intelligence. *Business & Information Systems Engineering*, 61: 637–643.

Denzau, A. T.; North, D. C.; et al. 1994. Shared mental models: ideologies and institutions. *KYKLOS-BERNE-*, 47: 3–3.

Dubey, A.; Abhinav, K.; Jain, S.; Arora, V.; and Puttaveerana, A. 2020. HACO: a framework for developing human-AI teaming. In *Proceedings of the 13th Innovations in Software Engineering Conference on Formerly known as India Software Engineering Conference*, 1–9.

Endsley, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1): 32–64.

Evertsz, R.; and Thangarajah, J. 2020. A Framework for Engineering Human/Agent Teaming Systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2477–2484.

Feng, Y.-H.; Teng, T.-H.; and Tan, A.-H. 2009. Modelling situation awareness for context-aware decision support. *Expert Systems with Applications*, 36(1): 455–463.

Gronauer, S.; and Diepold, K. 2022. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 1–49.

Gupta, S.; Hazra, R.; and Dukkipati, A. 2020. Networked multi-agent reinforcement learning with emergent communication. *arXiv preprint arXiv:2004.02780*.

Gupta, S.; Shilton, A.; AV, A. K.; Ryan, S.; Abdolshah, M.; Le, H.; Rana, S.; Berk, J.; Rashid, M.; and Venkatesh, S. 2023. BO-Muse: A human expert and AI teaming framework for accelerated experimental design. *arXiv preprint arXiv:2303.01684*.

Hani Daniel Zakaria, M.; Lengagne, S.; Corrales Ramón, J. A.; and Mezouar, Y. 2021. General framework for the optimization of the human-robot collaboration decision-making process through the ability to change performance metrics. *Frontiers in Robotics and AI*, 8: 736644.

Jackson, R. B.; and Williams, T. 2021. A theory of social agency for human-robot interaction. *Frontiers in Robotics and AI*, 8: 687726.

Johnson, C. J.; Demir, M.; McNeese, N. J.; Gorman, J. C.; Wolff, A. T.; and Cooke, N. J. 2021. The impact of training on human–autonomy team communications and trust calibration. *Human factors*, 00187208211047323.

Khamassi, M.; Velentzas, G.; Tsitsimis, T.; and Tzafestas, C. 2018. Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4): 881–893.

Khavas, Z. R.; Ahmadzadeh, S. R.; and Robinette, P. 2020. Modeling trust in human-robot interaction: A survey. In *Social Robotics: 12th International Conference, ICSR 2020, Golden, CO, USA, November 14–18, 2020, Proceedings 12*, 529–541. Springer.

Lakhmani, S.; Abich, J.; Barber, D.; and Chen, J. 2016. A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams. In *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience: 10th International Conference, AC 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17-22, 2016, Proceedings, Part II 10*, 296–307. Springer.

Lohani, M.; Stokes, C.; Dashan, N.; McCoy, M.; Bailey, C. A.; and Rivers, S. E. 2017. A framework for human-agent social systems: the role of non-technical factors in operation success. In *Advances in Human Factors in Robots and Unmanned Systems: Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems, July 27-31, 2016, Walt Disney World®, Florida, USA*, 137–148. Springer.

Lyons, J. B.; Sycara, K.; Lewis, M.; and Capiola, A. 2021. Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, 12: 589585.

Ma, W.; Chang, Y.-C.; Wang, Y.-K.; and Lin, C.-T. 2022. Human-Autonomous Teaming Framework Based on Trust Modelling. In *AI 2022: Advances in Artificial Intelligence: 35th Australasian Joint Conference, AI 2022, Perth, WA, Australia, December 5–8, 2022, Proceedings*, 707–718. Springer.

Martinez-Gil, F.; Lozano, M.; and Fernández, F. 2017. Emergent behaviors and scalability for multi-agent reinforcement learning-based pedestrian models. *Simulation Modelling Practice and Theory*, 74: 117–133.

Nothwang, W. D.; Gremillion, G. M.; Donavanik, D.; Haynes, B. A.; Atwater, C. S.; Canady, J. D.; Metcalfe, J. S.; and Marathe, A. R. 2016. Multi-sensor fusion architecture for human-autonomy teaming. In *2016 Resilience Week (RWS)*, 166–171. IEEE.

O'Neill, T.; McNeese, N.; Barron, A.; and Schelble, B. 2022. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors*, 64(5): 904–938.

Poole, D. L.; and Mackworth, A. K. 2010. *Artificial Intelligence: foundations of computational agents*. Cambridge University Press.

Richardson, R. C. 2017. Heuristics and satisficing. In Bechtel, W.; and Graham, G., eds., *A companion to cognitive science*, 566–575. Wiley Online Library.

Schelble, B.; Canonico, L.-B.; McNeese, N.; Carroll, J.; and Hird, C. 2020. Designing human-autonomy teaming experiments through reinforcement learning. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 64, 1426–1430. SAGE Publications Sage CA: Los Angeles, CA.

Schelble, B. G.; Flathmann, C.; and McNeese, N. 2020. Towards meaningfully integrating human-autonomy teaming in applied settings. In *Proceedings of the 8th international conference on human-agent interaction*, 149–156.

Scheutz, M.; DeLoach, S. A.; and Adams, J. A. 2017. A framework for developing and using shared mental models in human-agent teams. *Journal of Cognitive Engineering and Decision Making*, 11(3): 203–224.

Sharma, P. K.; Fernandez, R.; Zaroukian, E.; Dorothy, M.; Basak, A.; and Asher, D. E. 2021. Survey of recent multi-agent reinforcement learning algorithms utilizing centralized training. In *Artificial intelligence and machine learning for multi-domain operations applications III*, volume 11746, 665–676. SPIE.

Spaan, M. T. 2012. Partially observable Markov decision processes. *Reinforcement learning: State-of-the-art*, 387–414.

Sunehag, P.; Lever, G.; Gruslys, A.; Czarnecki, W. M.; Zambaldi, V.; Jaderberg, M.; Lanctot, M.; Sonnerat, N.; Leibo, J. Z.; Tuyls, K.; et al. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*.

Sutton, R. S.; Barto, A. G.; et al. 1998. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge.

Tai, J. E. M. 2021. *Coactive Design in Systems Engineering: Human-Machine Teaming in Search and Rescue (SAR) Operations*. Ph.D. thesis, Monterey, CA; Naval Postgraduate School.

Zambaldi, V.; Raposo, D.; Santoro, A.; Bapst, V.; Li, Y.; Babuschkin, I.; Tuyls, K.; Reichert, D.; Lillicrap, T.; Lockhart, E.; et al. 2019. Deep reinforcement learning with relational inductive biases. In *International conference on learning representations*.

Zhang, K.; Yang, Z.; and Başar, T. 2021a. Decentralized multi-agent reinforcement learning with networked agents: Recent advances. *Frontiers of Information Technology & Electronic Engineering*, 22(6): 802–814.

Zhang, K.; Yang, Z.; and Başar, T. 2021b. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, 321–384.