Project Summary

The science of autonomy is multi-faceted, presenting formidable open problems. Significant progress has been made on short-term autonomy for physical and software agents that sense and interact with the environment and with teammates. However, long(er)-duration autonomy in a team of heterogeneous agents operating without prior coordination in complex, chaotic, resource-constrained, and potentially adversarial environments, remains an open problem. As a significant step towards addressing this problem, we seek to jointly address the key underlying knowledge representation and reasoning, learning, and multiagent collaboration challenges. We will do so by developing algorithms and architectures for the following research thrusts:

- **Explainable reasoning and learning:** we will develop a novel architecture that tightly couples the principles and complementary strengths of non-monotonic logical reasoning, probabilistic reasoning, data-driven statistical learning, and formal models. This architecture will enable agents to (a) represent, reason with, and learn from multi-level, incomplete, commonsense domain knowledge and observations; (b) interactively provide explanatory descriptions of decisions, beliefs, and experiences at a suitable level of abstraction; and (c) establish that their behavior satisfies certain desirable properties. We will ground and evaluate this architecture in the challenging context of ad hoc multiagent teamwork.
- **Communication in ad hoc teamwork:** a common assumption in ad-hoc teamwork is the absence of communication between agents during task execution. However, once agents start their task in many real world situations, even if they do not know the goals or plans of other agents, they can communicate using network protocols, controlled vocabularies, or visual and audio signals. We propose to investigate such scenarios by (a) defining a taxonomy of possible communication types for agents in ad hoc teamwork, especially when agents have to perform a complex multi-step task; (b) quantifying the value of communication and deciding when to ask a question and what to ask, thus supporting decentralized reasoning; and (c) proposing new algorithms and a test-bed for evaluation for novel agent designs.
- **Open ad hoc teamwork:** we will significantly extend ad hoc teamwork to *open* teams in which agents may enter or leave at any time and without prior announcement or opportunities to calibrate teamwork strategies. Specifically, we will enable: (a) the use of graph neural networks (GNNs) for reasoning and learning about dynamic team composition and agent relations in open ad hoc teamwork; (b) integration of GNNs with deep reinforcement learning methods to generate optimal control policies for open ad hoc teamwork under variable observability conditions; and (c) enable joint training of multiple agents to collaborate more efficiently in open ad hoc teamwork settings.

These research thrusts are grounded in principled, theoretically-motivated methods, with a strong emphasis on empirical evaluation in complex, multiagent domains. Evaluation scenarios will consist of collaborative deployment of agents or robots with humans in simulation and in the real world. These thrusts are related to multiple ONR-wide hard problems. They concentrate on “embodied and situated intelligence and architectures” and on “scalable and robust distributed collaboration”, but also touch upon aspects of focus areas such as “human/autonomy collaboration and teaming” and “robust and trustworthy autonomy”.