Grounded Communication Policies in Heterogeneous Agent Reinforcement Learning

Doctoral Consortium

Aju Ani Justus
University of Birmingham
Birmingham, United Kingdom
axa1943@student.bham.ac.uk

ABSTRACT

Communication allows agents to overcome individual perceptual limitations, coordinate their behaviours, and operate collectively in multi-agent systems. Most existing approaches treat communication as an information-passing channel embedded within agents' action policies. As a result, communication is often abstracted as information exchange, whereas communication in humans typically involves establishing and maintaining common ground, that is, shared understanding of task-relevant information. This consideration becomes particularly relevant in heterogeneous-agent reinforcement learning (HARL), where agents differ in their perceptual or action capabilities. In this paper, I argue that communication in HARL should be modelled as a distinct, learnable process aimed at grounding shared beliefs, rather than as an auxiliary component of the action policy. I propose a framework in which each agent learns a separate communication policy that operates in parallel with its action policy to iteratively establish common ground before and during task execution. My work investigates how such grounded communication policies affect coordination, learning stability, and generalisation in HARL.

KEYWORDS

multi-agent systems; agent communication; multi-agent reinforcement learning; heterogeneous agents; coordination

ACM Reference Format:
Aju Ani Justus. 2026. Grounded Communication Policies in Heterogeneous Agent Reinforcement Learning: Doctoral Consortium. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/OAYF2008

1 INTRODUCTION AND MOTIVATION

Multi-agent reinforcement learning (MARL) studies how multiple decision-making agents learn to interact with an environment and with one another. In many domains of interest, such as robot teams [6] and autonomous vehicles [15], agents operate under partial observability and must coordinate to achieve shared or overlapping goals. As MARL has evolved toward decentralized and adaptive systems tackling increasingly complex tasks, a critical avenue for further research has emerged: meaningful communication. Most existing MARL algorithms that use communication for learning tasks treat communication as information exchange, often using encoded observations or hidden states as messages to broaden an agent's view of the environment. However, as systems scale to encompass agents with diverse sensors, actuators, and objectives, communication should be modelled as a distinct, learnable process aimed at grounding shared beliefs, rather than as an auxiliary component of the action policy.

A large body of recent work has demonstrated that learned communication can significantly improve performance, spanning from small-scale games to complex video games. In early work, Foerster *et al.* [3] developed two simple games, namely *Switch Riddle* and *MNIST Games*, to evaluate their proposed models, DIAL and RIAL. Sukhbaatar *et al.* [18] introduced the *Traffic Junction* environment to evaluate CommNet, which has since become a popular benchmark in subsequent studies [2, 5, 7, 11, 12, 16]. Among these works, MAGIC [12] achieved superior performance on *Traffic Junction* with local rewards when compared to earlier approaches such as CommNet [18], IC3Net [16], and the more recent GA-Comm [11]. StarCraft and its SMAC variant are widely used benchmarks for cooperative MARL [14, 19, 20]. Numerous communication-based approaches have been evaluated in this setting [7, 13, 21–26]. Recent approaches such as FCMNet [23] and MAIC outperform multiple communication-based and value decomposition methods (e.g., QMIX) on various maps. To date, only MAGIC [12] has reported results on Google Research Football [10] using communication. IC3Net [16], TarMAC [2], and MAGIC [12] are evaluated on mixed predator-prey environments, where agents learn to communicate only when necessary. NDQ [22] is studied in an independent search scenario and shows that agents learn not to communicate when goals are independent.

Despite this progress, the dominant modelling assumption in communication-based MARL remains surprisingly simplistic: messages are treated as just another input to a policy network or as part of the action space. In most architectures, the same policy that selects environment actions also implicitly decides what information to communicate, with no explicit representation of whether the information is understood, aligned, or shared among agents. This assumption is particularly limiting in heterogeneous-agent reinforcement learning (HARL), where agents differ in their observation modalities, perceptual resolution, or action capabilities. In such settings, coordination requires more than broadcasting observations, it requires agents to form a mutual understanding of how information should be interpreted. For example, if one agent can perceive colour and another can perceive shape, coordinating on a target object

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/OAYF2008

requires establishing which attributes refer to the same entity. This challenge is conceptually distinct from works such as IC [4], which studies heterogeneous-agent settings which enables effective collaboration by transferring information from better-informed agents to better-acting agents, but it does not explicitly address how agents establish or negotiate shared semantic interpretations of messages. As such, the meaning of communication in IC is implicitly aligned through training rather than grounded through an explicit process of common-ground formation.

The concept of common ground is deeply rooted in the study of human language use, which is viewed not as a simple transmission of information but as a form of joint action [1]. Human communication research of dialogue emphasizes that joint action relies on common ground, defined as the shared knowledge and assumptions that interlocutors believe they share. Communication is not a one-shot transmission of symbols but an iterative process involving clarification, confirmation, and repair [8]. Inspired by this view, my research asks: How can HARL systems explicitly learn to establish common ground? Related ideas appear in adjacent literatures on grounded and emergent communication, most notably in robotics and language emergence, where shared meaning arises through repeated interaction [17] and in recent MARL work, similarly inspired by human language, that investigates the emergence of recursive communication via bootstrapping and iterated learning [9].

In this work, common ground is not the same as a richer environment state or an agent's observation space, but instead refers to a dynamic and potentially incomplete set of task-relevant beliefs that agents assume to be mutually held and that may require ongoing maintenance, revision, or repair through interaction. Explicitly modelling such grounded common ground is likely to incur additional computational and communication overhead and raises a number of open challenges, including deciding what information should persist over time, how relevance is determined, and how misalignment is detected and corrected. This work treats these issues as central research questions for grounded communication in HARL and hypothesizes that addressing them is necessary for scaling HARL beyond simple information exchange toward domains requiring semantic alignment.

2 PROPOSED APPROACH

I propose a framework for grounded communication in HARL based on decoupling communication from action.

2.1 Decoupled Communication and Action Policies

Each agent is equipped with two distinct learned components:

(1) Communication Policy: This policy determines when to communicate, what messages to send, and how to respond to received messages. Its objective is to update a shared belief or latent context representing common ground among agents. It will be parameterised as a recurrent neural network (e.g., LSTM or GRU) that maintains an internal dialogue state across communication turns.
(2) Action Policy: This policy selects environment actions based on the agent's local observation and the grounded information provided by the communication process.

Communication occurs in parallel with action selection and may involve multiple turns per environment step, allowing agents to iteratively clarify and align their beliefs. Crucially, the communication policy is not intended to transmit all available information, but to select messages that are expected to be relevant to other agents' decision-making, conditioned on inferred beliefs about their capabilities and likely actions.

2.2 Representing Common Ground

Common ground is represented as a recurrent latent memory, updated through communication actions using gated recurrent mechanisms. This representation is not assumed to be globally observable but is approximated through message exchanges. The communication policy is trained to reduce uncertainty or disagreement about task-relevant variables within this shared context.

2.3 Learning Objectives

In addition to task rewards, the communication policy may be trained with auxiliary objectives that encourage grounding, such as: agreement or consistency between agents' inferred beliefs, mutual information between communicated messages and relevant latent variables, importantly, penalties for redundant or unnecessary communication. These objectives are designed to shape communication behaviour without interfering directly with action learning.

3 PLANNED RESEARCH

The planned research will proceed in three stages. First, I will implement the proposed decoupled communication and action framework in a set of controlled heterogeneous-agent environments, beginning with grid-world and predator–prey tasks. These environments allow systematic manipulation of heterogeneity and provide interpretable settings for studying grounding behaviour.

Second, I will evaluate the effects of grounded communication on coordination, learning stability, and sample efficiency by comparing against established communication-based MARL baselines that embed communication within action policies. Ablation studies will examine the role of recurrent communication memory and multi-turn communication in establishing common ground.

Finally, I will investigate generalisation by transferring learned communication policies to novel tasks, unseen agent combinations, or altered observation modalities. This will assess whether explicitly grounding communication enables more robust coordination under distributional shift compared to latent-message approaches.

4 SIGNIFICANCE

By modelling communication as a distinct, learnable process aimed at grounding common beliefs, this work moves beyond treating messages as latent information vectors and focuses on how meaning is formed and maintained. This approach has implications for scaling multi-agent systems to real-world domains where agents differ in sensing or capabilities, such as mixed robot teams or human–AI teaming. While grounded communication increases computational and communication overhead, embracing this complexity is essential for enabling coordination in tasks requiring semantic alignment, expanding the class of problems HARL can reliably solve.

ACKNOWLEDGMENTS

Thanks to Chris Baber and Leonardo Stella for the support and the cooperative agreement award (W911NF-22-2-0161) from the DEVCOM Army Research Laboratory to the Alan Turing Institute and University of Birmingham.

REFERENCES

[1] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition*, Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley (Eds.). American Psychological Association, 127–149.

[2] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Mike Rabbat, and Joelle Pineau. 2019. TarMAC: Targeted Multi-Agent Communication. In *International Conference on Machine Learning*. Proceedings of Machine Learning Research, 1538–1546.

[3] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with Deep multi-agent reinforcement learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Barcelona, Spain) *(NIPS'16)*. Curran Associates Inc., Red Hook, NY, USA, 2145–2153.

[4] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gülçehre, Pedro A. Ortega, DJ Strouse, Joel Z. Leibo, and Nando de Freitas. 2019. Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. In *International Conference on Machine Learning*. Proceedings of Machine Learning Research, 3040–3049.

[5] Woojun Kim, Jongeui Park, and Youngchul Sung. 2021. Communication in Multi-Agent Reinforcement Learning: Intention Sharing. In *International Conference on Learning Representations*. https://openreview.net/forum?id=qpsl2dR9twy

[6] Jens Kober, J. Andrew Bagnell, and Jan Peters. 2013. Reinforcement Learning in Robotics: A Survey. *International Journal of Robotics Research* 32, 11 (2013), 1238–1274. https://doi.org/10.1177/0278364913495721

[7] Xiangyu Kong, Bo Xin, Fangchen Liu, and Yizhou Wang. 2017. Revisiting the Master-Slave Architecture in Multi-Agent Deep Reinforcement Learning. *CoRR* abs/1712.07305 (2017).

[8] Theodora Koulouri, Stanislao Lauria, and R. D. Macredie. 2016. Do (and say) as I say: Linguistic adaptation in human-computer dialogs. *Human-Computer Interaction* 31, 1 (2016), 59–95. https://doi.org/10.1080/07370024.2014.934180

[9] Vikas Kumar and Ajin George Joseph. 2025. Emergence of Recursive Language through Bootstrapping and Iterated Learning. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025) (IFAAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, Detroit, Michigan, USA, 1235–1243.

[10] Karol Kurach et al. 2020. Google Research Football: A Novel Reinforcement Learning Environment. In *AAAI Conference on Artificial Intelligence*. AAAI Press, 4501–4510.

[11] Yong Liu, Weixun Wang, Yujing Hu, Jianye Hao, Xingguo Chen, and Yang Gao. 2020. Multi-Agent Game Abstraction via Graph Attention Neural Network. In *AAAI Conference on Artificial Intelligence*. AAAI Press, 7211–7218.

[12] Yaru Niu, Rohan R. Paleja, and Matthew C. Gombolay. 2021. Multi-Agent Graph-Attention Communication and Teaming. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 964–973.

[13] Peng Peng, Quan Yuan, Ying Wen, Yaodong Yang, Zhenkun Tang, Haitao Long, and Jun Wang. 2017. Multiagent Bidirectionally-Coordinated Nets for Learning to Play StarCraft Combat Games. *CoRR* abs/1703.10069 (2017).

[14] Mikayel Samvelyan, Tabish Rashid, Christian Schröder de Witt, et al. 2019. The StarCraft Multi-Agent Challenge. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2186–2188.

[15] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. 2016. Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving. *CoRR* abs/1610.03295 (2016).

[16] Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. 2019. Learning When to Communicate at Scale in Multiagent Cooperative and Competitive Tasks. In *International Conference on Learning Representations*.

[17] Luc Steels. 2003. Evolving grounded communication for robots. *Trends in Cognitive Sciences* 7, 7 (2003), 308–312. https://doi.org/10.1016/S1364-6613(03)00129-3

[18] Sainbayar Sukhbaatar, Arthur Szlam, and Rob Fergus. 2016. Learning Multiagent Communication with Backpropagation. In *Advances in Neural Information Processing Systems*, Vol. 29. Curran Associates Inc., 2244–2252.

[19] Gabriel Synnaeve, Nantas Nardelli, Alex Auvolat, Soumith Chintala, Timothée Lacroix, Zeming Lin, Florian Richoux, and Nicolas Usunier. 2016. TorchCraft: A Library for Machine Learning Research on Real-Time Strategy Games. *CoRR* abs/1611.00625 (2016).

[20] Oriol Vinyals et al. 2017. StarCraft II: A New Challenge for Reinforcement Learning. *CoRR* abs/1708.04782 (2017).

[21] Rundong Wang, Xu He, Runsheng Yu, Wei Qiu, Bo An, and Zinovi Rabinovich. 2020. Learning Efficient Multi-Agent Communication: An Information Bottleneck Approach. In *International Conference on Machine Learning*. Proceedings of Machine Learning Research, 9908–9918.

[22] Tonghan Wang, Jianhao Wang, Chongyi Zheng, and Chongjie Zhang. 2020. Learning Nearly Decomposable Value Functions via Communication Minimization. In *International Conference on Learning Representations*.

[23] Yutong Wang and Guillaume Sartoretti. 2022. FCMNet: Full Communication Memory Net for Team-Level Cooperation in Multi-Agent Systems. *CoRR* abs/2201.11994 (2022).

[24] Lei Yuan, Jianhao Wang, Fuxiang Zhang, Chenghe Wang, Zongzhang Zhang, Yang Yu, and Chongjie Zhang. 2022. Multi-Agent Incentive Communication via Decentralized Teammate Modeling. In *AAAI Conference on Artificial Intelligence*. AAAI Press.

[25] Sai Qian Zhang, Qi Zhang, and Jieyu Lin. 2019. Efficient Communication in Multi-Agent Reinforcement Learning via Variance Based Control. In *Advances in Neural Information Processing Systems*, Vol. 32. Curran Associates Inc., 3230–3239.

[26] Sai Qian Zhang, Qi Zhang, and Jieyu Lin. 2020. Succinct and Robust Multi-Agent Communication with Temporal Message Control. In *Advances in Neural Information Processing Systems*, Vol. 33. Curran Associates Inc.