Measuring cooperation among competing AI algorithms: A framework based on empirical game-theoretic analysis

1 Problem Statement

Recent months have witnessed the advancement of surprisingly capable AI at an unprecedented pace and its deployment—whether actualized, proposed, or in-the-works—in a wide range of application domains. One natural consequence is that AI artifacts may be expected to perform in complex environments that differ significantly from the contexts in which they were trained (or which the original developers had in mind), including arbitrary, complex multiagent environments. This poses novel challenges for *evaluating* the safety and beneficence of advanced AI, particularly regarding properties that emerge from the *interactions* of autonomous agents, both human and artificial.

The imperative to evaluate advanced AI is widely appreciated, whether for performance, safety assessment, or other design objective. For tracking progress in LLM-based agents, many have turned to Elo ratings [Stern, 2024], an approach designed for zero-sum games. This measure is quite narrow and ill-suited for the vast majority of contexts where agent interactions are not primarily adversarial. (The trick of using Elo is to define the evaluation as a zero-sum competition among alternatives.) We seek instead an approach that applies to a broad range of scenarios where agents meaningfully interact but are not adversarial, nor can their interests be assumed to be perfectly aligned. The approach should be agnostic to technology, focusing on the outcomes of the agent interactions. It should also support a broad palette of possible metrics, answering different questions that evaluators may care about, accounting for uncertainty in these answers.

An example domain that we will explore in detail is that of agent-based negotiation. This could involve AIs negotiating with humans, or on behalf of humans with other AIs or humans. In the most relevant instances, negotiation is not zero-sum. Indeed, much of the art of negotiation is in constructing agreements that allow both parties to satisfy their objectives to a large extent. At the same time, these objectives are in some degree of tension, so there are inherent tradeoffs between pursuit of self-interest and common goals. Negotiation scenarios can range from bargaining over terms in a well-defined parameter space to mutual development of a plan to work on a collaborative project.

The goal of our work is to evaluate the cooperative capabilities and related properties of advanced AI methods operating in such domains. One key issue we must address is that any measure used for evaluation has the potential to distort behavior. (Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure".) Ironically, rewarding cooperation per se could have unanticipated non-cooperative effects, such as encouraging performative cooperation. We will build on prior work that addressed the issue of incentives for cooperation in a competitive environment [Sinha and Wellman, 2019].

We propose to develop in this project a novel, principled framework for the evaluation of advanced, interactive AI algorithms in general multiagent environments, focusing attention on those that inherently contain cooperative and competitive elements. Our approach is based on the recent development of a *meta-game* approach to evaluating multiagent training algorithms [Li and Wellman, 2024], of the sort employed in multiagent reinforcement learning. It provides for quantification of uncertainty about evaluations, using statistical bootstrapping techniques. Our project comprises three broad tasks:

(i) We will develop suitable metrics of AI cooperation, addressing incentive issues [Sinha and Wellman, 2019]. While striving for generality, we will also consider how to take advantage of special structure that may be exhibited in environment classes of interest.

(ii) We will incorporate these metrics in our meta-game evaluation approach. The techniques will be extended to address algorithms that train agents based on large language models.

(iii) We will demonstrate the efficacy of our approach in negotiation scenarios, starting with alternating-offer bargaining for the allocation of a set of indivisible goods [Fatima et al., 2014, DeVault et al., 2015, Lewis et al., 2017] and proceeding to more complex collaborative planning.

2 Background

Our starting reference point is prior work on evaluation protocols for ***multiagent reinforcement learning*** (`MARL`) algorithms. Much of this has focused on purely cooperative settings [Papoudakis et al., 2020, Gorsane et al., 2022], where global team reward is clearly defined as the common objective [Foerster et al., 2018, Rashid et al., 2020]. In such environments, cooperation is a given, but even so, complex learning dynamics may lead agents using the same algorithm to coordinate on distinct machine conventions in different runs [Hu et al., 2020, Bakhtin et al., 2021]. Game-theoretic analysis of purely cooperative settings often focuses on formation of coalitions and the stability of coalition agreements [Chalkiadakis et al., 2022].

At the other extreme, purely adversarial (*i.e.*, two-player zero-sum) environments, `MARL` algorithms can be evaluated in terms of exploitability (equivalently, distance to Nash equilibrium) [Brown et al., 2020, Schmid et al., 2023], as all equilibria are interchangeably optimal. In such settings, cooperation is strictly speaking impossible.

More generally, where there are multiple equilibria or where we do not necessarily expect equilibrium behavior, the metrics for `MARL` performance may be less clear. While evaluating single-agent (deep) reinforcement learning algorithms is well-studied [Henderson et al., 2018, Jordan et al., 2020, Agarwal et al., 2021], there are relatively few works on `MARL` evaluation principles [Gronauer and Diepold, 2022]. Typical evaluation protocols measure performance against a selected set of background opponents or context-specific emergent behaviors [Lowe et al., 2017, Li et al., 2019, Song et al., 2020, Leibo et al., 2021]. Evaluating performance against humans has been one source of compelling demonstrations [Silver et al., 2016, Vinyals et al., 2019, Wurman et al., 2022, Perolat et al., 2022], but this approach is limited by the range of tasks for which human expertise exists, and the cost of engaging it when it is available.

Game Theory Preliminaries. A ***game*** $\mathcal{G}$ is a formal representation of a multiagent strategic scenario, specifying all *strategies* available to each agent or *player* in the scenario of interest as well as the *utility* of each agent as a function of the *strategy profile*, that is, the combination of all agents' strategies. The representation can assume multiple *forms* depending on which salient aspects of strategic interaction we wish to encode. A ***normal-form*** game treats each strategy choice as a discrete, atomic object (a *pure* strategy) and assigns a vector of agent utility values to each pure strategy profile. A ***mixed strategy*** for player i defines a probability distribution over that player's pure strategy space; player i's ***payoff*** for a mixed strategy profile is given by its expected utility with respect to the joint distribution induced by the profile. The ***regret*** of player for a profile is the maximum improvement in payoff it can achieve by deviating to a pure strategy when all other players stick to their respective strategies in the profile. A ***Nash equilibrium*** (NE) is a strategy profile for which no player has positive regret. A game is ***symmetric*** if all players share the same strategy space and the same utility function, and each player's utility is permutation-invariant to other players' strategies; a profile is symmetric if every player adopts the same (mixed) strategy.

ETGA. Many realistic multiagent scenarios are so complex that an explicit *declarative* expression even in the normal form is infeasible, and at best only a *procedural* description may be available in the form of a high-fidelity *black-box simulator*, making it a *simulation-based game* [Vorobeychik and Wellman, 2008]. ***Empirical game-theoretic analysis*** **(EGTA)** is a general methodology for reasoning about games through such agent-based simulation [Tuyls et al., 2020, Wellman, 2016] by estimating a normal-form ***empirical game*** model for the intractable underlying multiagent scenario. This model is defined over an enumerated strategy set, a small selection from the full underlying strategy space, and is estimated using payoff data samples obtained by feeding selected strategy profiles into the simulator. A natural way

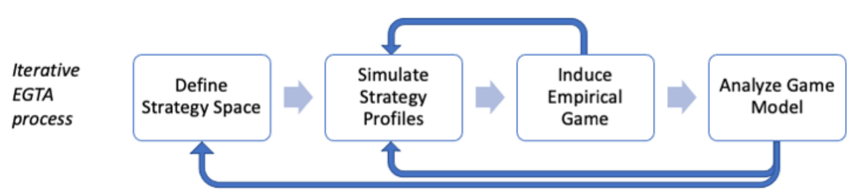

Figure 1: Schematic illustration of the EGTA framework.

to implement EGTA is to proceed iteratively (see Fig. 1), accumulating strategies and simulation data to continually expand and update the empirical game model. The trick is to trade off fidelity against tractability so that it is feasible to apply traditional game-solving techniques to the empirical model and yet obtain a solution to this empirical model that tells us something meaningful about the underlying game. *Policy space response oracles* (PSRO), introduced by DeepMind [Lanctot et al., 2017] and extended by many others [Bighashdel et al., 2024], accomplishes the key EGTA operation of extending agent strategy spaces with new strategies by first applying a chosen solution concept to the empirical game, then using `dRL` to compute each agent's true-game best response to this solution, and then augmenting the model with these best responses.

3 Proposed Methodology

3.1 Metrics of Cooperativeness

Our central goal is to devise metrics and procedures to evaluate the *degree of cooperativeness* exhibited by an arbitrary AI algorithm in an arbitrary multiagent environment that lies somewhere on the spectrum between purely adversarial and purely cooperative. We view cooperation as an *emergent property* of strategic interactions. That is, we quantify cooperativeness based on outcomes rather than on the form of interaction. Sinha and Wellman [2019] considered a related question: how to measure how *collaborative* an agent has been in the course of a competitive interaction. The issue is how to reward social benefit and at the same time capture the agent's individual contribution to that benefit. In other words, cooperativeness is not just a matter of regarding others' welfare; it is also about creativity and competence in promoting it.

Rewarding global social welfare aligns incentives for cooperation, but fails to distinguish individual contributions. With this in mind, Sinha and Wellman [2019] proposed a measure called ***local social welfare maximization*** (LSWM), which quantifies the utility accrued by the agent and its neighbors locally according to an evolving collaboration graph. LSWM was shown to exhibit desirable properties both theoretically and experimentally under various assumptions on the utility structure. However, the approach has limitations: scoring requires the knowledge of the evolving collaboration structure and certain agent-specific parameters as well as the ability to abstract agent action into a vector of fractions each representing the amount of collaboration effort extended to a network neighbor. Taking LSWM as a starting point, we will generalize the approach and develop a suite of metrics that will allow us to assess cooperativeness within a broad class of multiagent settings.

3.2 Meta-Game Evaluation Framework

Any metric for assessing an agent's cooperativeness will be sensitive to *strategic context*: assumptions about other-agent behavior. That is, whether an agent is good at cooperating will generally hinge on what the other agents do. Therefore, an overall evaluation scheme needs a principled way to decide the strategic context in which to evaluate the agents. For this purpose, we will adapt a general-purpose *meta-game* framework recently introduced by Li and Wellman [2024].

Formally, the subject of analysis is a ***multi-agent training algorithm*** (MATA) $\mathcal{M}$, a stochastic procedure that produces a joint policy (in `MARL` language) or strategy profile (in game-theoretic language) $\pi = \mathcal{M}(\mathcal{G}, \Theta, \omega)$. The first input is a game $\mathcal{G}$ provided in the form of a simulator only (e.g., a negotiation environment for LLMs). Θ is the set of hyperparameters of the MATA, specifying the neural architecture, learning rate, etc. The final input ω is a random seed, which accounts for the known uncertainty of a MATA in strategy generation. For example, Hu et al. [2020] and Bakhtin et al. [2021] have observed that different runs of a MATA with the same $\mathcal{G}$ and Θ but different ω may generate policies with vastly different strategic behaviors (e.g., strategies that adopt distinct offering conventions in a negotiation game). If $\mathcal{G}$ is symmetric, we can simplify analysis by focusing on a symmetric output π (*i.e.*, single policy to be played by all). We assume that the hyperparameters Θ^m for each MATA $\mathcal{M}^m$ have been fixed, so the uncertainty in behavior of a training algorithm is fully captured by the random seeds. An ordered pair of MATA and associated hyperparameters $(\mathcal{M}, \Theta)$ is called a ***parametrized MATA***.

The ***meta-game*** evaluation framework analyzes the relative performance of M parametrized MATAs by framing their interaction itself as a normal-form game across different *combinations* of seeds; a MATA itself acts as a *meta-strategy* in this framework since it generates a strategy profile given a game (among other inputs). Given a

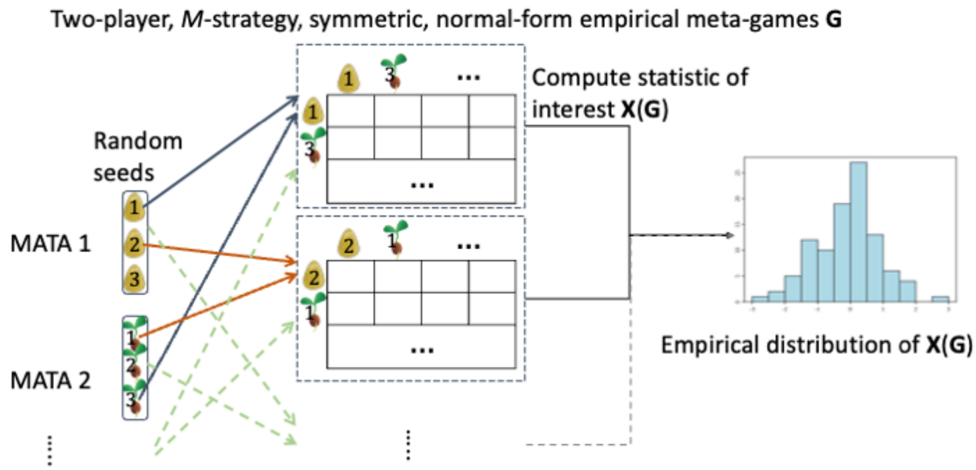

Figure 2: Schematic illustration of the meta-game evaluation procedure for two-player MATAs.

game $\mathcal{G}$, parametrized MATAs $\{(\mathcal{M}^1, \Theta^1), \ldots, (\mathcal{M}^M, \Theta^M)\}$, and arbitrary statistics $\boldsymbol{X}$ characterizing the strategic properties of interest, the meta-game evaluation procedure consists of the following steps:

1. Select a finite set of seeds Ω^m for each MATA m. Generate joint policy $\hat{\pi}^m(\omega) = \mathcal{M}^m(\mathcal{G}, \Theta^m, \omega)$ for each $\omega \in \Omega^m$.
2. For each m, uniformly sample $|\Omega^m|$ seeds from Ω^m with replacement, yielding the sequence $(\omega_j)_{j=1}^{|\Omega^m|}$. Let $\hat{\pi}^m$ be a profile that is payoff-equivalent to a uniform mixture over the multiset $\{\hat{\pi}^m(\omega_j)\}_{j=1}^{|\Omega^m|}$.
3. Given $\hat{\Pi} = \{\hat{\pi}^1, \ \ldots, \ \hat{\pi}^M\}$, estimate the symmetric empirical meta-game $\mathcal{MG}(\hat{\Pi})$.
4. Compute the statistics-of-interest $\boldsymbol{X}$ from $\mathcal{MG}(\hat{\Pi})$. This may involve first solving $\mathcal{MG}(\hat{\Pi})$ according to some solution concept.
5. Repeat Steps 2 through 4. Memoize estimated profile payoffs for reuse across iterations. Obtain an empirical distribution of $\boldsymbol{X}$ and report statistical properties of $\boldsymbol{X}$.

EGTA comes into play in steps 3 and 4. By resampling from the set of seeds, Step 2 essentially accomplishes bootstrapping [Wiedenbeck et al., 2014], allowing us to construct sampling distributions over $\boldsymbol{X}$ from analyzing multiple empirical games among MATAs. Figure 2 presents a two-player version of this procedure.

For their proof-of-concept application of this framework, Li and Wellman [2024] studied a two-player game of alternating negotiation for indivisible item allocation under subjective valuations [Lewis et al., 2017]. They evaluated a suite of state-of-the-art `MARL` algorithms including variants of AlphaZero [Silver et al., 2018], Multi-agent Proximal Policy Optimization (MAPPO) [Yu et al., 2022], Regularized Nash Dynamics (R-NaD) [Perolat et al., 2021], and Neural Fictitious Self-Play (NFSP) [Heinrich and Silver, 2016]. One of the statistics they recorded from a meta-game was the frequency with which each of these MATAs turned out to be a best response (i.e., payoff-maximizer) to every other MATA and itself. These results, visualized in Fig. 3, reveal interesting insights, for example that most of these MATAs tend to overfit to interactions with themselves.

3.3 Cooperativeness Testbed

Putting things together, we can evaluate agents' cooperativeness using this meta-game framework, by employing suitable cooperativeness metrics for the statistics $\boldsymbol{X}$. Note that in Step 4 of the meta-game evaluation procedure in Section 3.2, the choice of the solution concept is left open. For example, the original study employed max-entropy Nash equilibrium, in order to compute statistics $\boldsymbol{X}$ based on NE-regret or Nash averaging [Balduzzi et al., 2018,

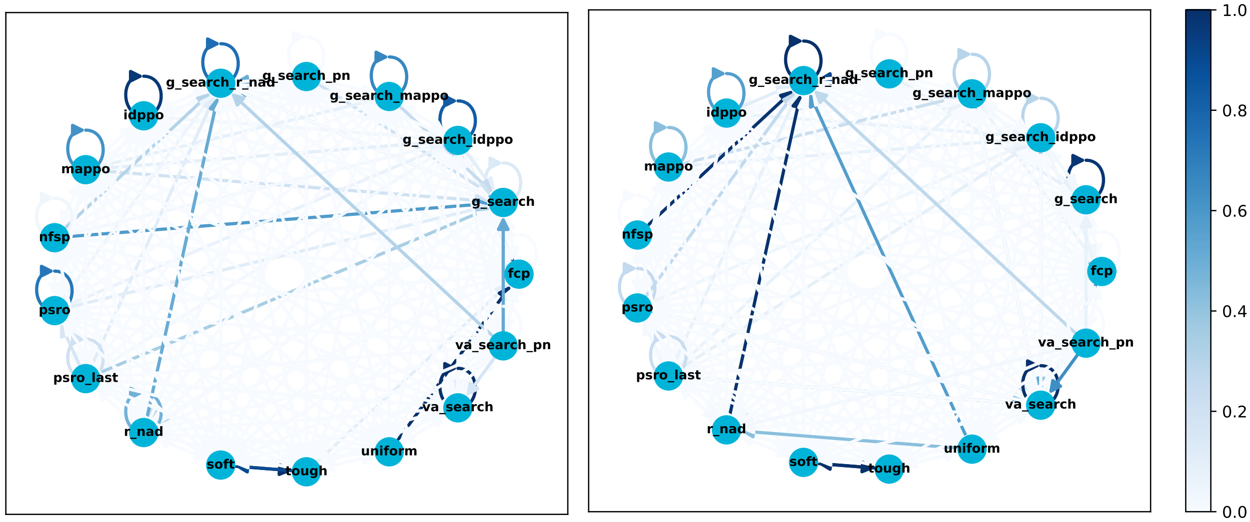

Figure 3: Empirical best-response graph for MATAs in a negotiation game generated using the meta-game framework [Li and Wellman, 2024]. Nodes represents MATAs. Darkness of a directed edge from a to b signifies the relative frequency of a being a best response to b in sampled empirical games. Note the lack of a dominant MATA, and the prevalence of self-edges.

Jordan et al., 2007]. A task in our proposed project will be to identify alternative solution concepts that may be more appropriate for assessing cooperation effectiveness.

We will implement and share an open-source meta-game package with support for a suite of cooperativeness metrics and interfaces for arbitrary advanced algorithms (e.g., deep MARL, LLMs) and game simulators. We will report on the application of this framework to select multiagent domains, in particular an augmentation of the negotiation scenario introduced in Section 3.2 with outside offers for each agent. We will also develop in-house simulators for these domains as per need — these by-products may be useful in their own right.

4 Outcomes, Significance, and Impact

Our goal is to rigorize, standardize and expedite the task of evaluating cooperativeness of new and emerging AI prior to fielded applications. This has strong implications for anticipating the behavior of these complex algorithms and systems, which can feed back into their design and fine-tuning. Moreover, we believe that, in the long term, the project can provide a basis for evaluators (including prospective users, regulators, and other third parties) for assessing the impact of new algorithms in the growing ecosystem of MATAs. We will seek to publish our findings at leading AI venues such as AAAI, IJCAI, AAMAS, NeurIPS, etc.

Potential Risks and Limitations. We must first acknowledge the potential risks naturally engendered by using any evaluation in a consequential context like regulation, by government or other authorities. On a more technical front, since our proposed evaluation framework is based on game theory and statistical techniques (namely, bootstrapping), it inherits the known drawbacks and limitations of these disciplines. Examples include the issues of non-unique equilibria and incorrect assumptions about other parties in game-theoretic analysis, and sampling issues affecting our statistical components. We plan to identify and address specific instances of these issues over the course of the project.

References

R. Agarwal, M. Schwarzer, P. S. Castro, A. C. Courville, and M. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *35th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 29304–29320, 2021.

A. Bakhtin, D. Wu, A. Lerer, and N. Brown. No-press diplomacy from scratch. In *35th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 18063–18074, 2021.

D. Balduzzi, K. Tuyls, J. Perolat, and T. Graepel. Re-evaluating evaluation. In *32nd International Conference on Neural Information Processing Systems (NeurIPS)*, 2018.

A. Bighashdel, Y. Wang, S. McAleer, R. Savani, and F. A. Oliehoek. Policy space response oracles: A survey. In *33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024.

N. Brown, A. Bakhtin, A. Lerer, and Q. Gong. Combining deep reinforcement learning and search for imperfect-information games. In *34th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 17057–17069, 2020.

G. Chalkiadakis, E. Elkind, and M. Wooldridge. *Computational aspects of cooperative game theory*. Springer International Publishing, 2022.

D. DeVault, J. Mell, and J. Gratch. Toward natural turn-taking in a virtual human negotiation agent. In *AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction*, 2015.

S. Fatima, S. Kraus, and M. Wooldridge. *Principles of Automated Negotiation*. Cambridge University Press, 2014.

J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson. Counterfactual multi-agent policy gradients. In *32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 2974–2982, 2018.

R. Gorsane, O. Mahjoub, R. J. de Kock, R. Dubb, S. Singh, and A. Pretorius. Towards a standardised performance evaluation protocol for cooperative MARL. In *36th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 5510–5521, 2022.

S. Gronauer and K. Diepold. Multi-agent deep reinforcement learning: A survey. *Artificial Intelligence Review*, 55:895–943, 2022.

J. Heinrich and D. Silver. Deep reinforcement learning from self-play in imperfect-information games. *arXiv preprint arXiv:1603.01121*, 2016.

P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep reinforcement learning that matters. In *32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 3207–3214, 2018.

H. Hu, A. Lerer, A. Peysakhovich, and J. Foerster. "Other-play" for zero-shot coordination. In *37th International Conference on Machine Learning (ICML)*, pages 4399–4410, 2020.

P. R. Jordan, C. Kiekintveld, and M. P. Wellman. Empirical game-theoretic analysis of the TAC supply chain game. In *6th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1188–1195, 2007.

S. Jordan, Y. Chandak, D. Cohen, M. Zhang, and P. Thomas. Evaluating the performance of reinforcement learning algorithms. In *37th International Conference on Machine Learning (ICML)*, pages 4962–4973, 2020.

M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *31st International Conference on Neural Information Processing Systems (NeurIPS)*, pages 4190–4203, 2017.

J. Z. Leibo, E. A. Dueñez-Guzman, A. Vezhnevets, J. P. Agapiou, P. Sunehag, R. Koster, J. Matyas, C. Beattie, I. Mordatch, and T. Graepel. Scalable evaluation of multi-agent reinforcement learning with melting pot. In *38th International Conference on Machine Learning (ICML)*, pages 6187–6199, 2021.

M. Lewis, D. Yarats, Y. Dauphin, D. Parikh, and D. Batra. Deal or No Deal? End-to-End Learning of Negotiation Dialogues. In *2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453, 2017.

S. Li, Y. Wu, X. Cui, H. Dong, F. Fang, and S. Russell. Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient. In *33rd AAAI Conference on Artificial Intelligence*, pages 4213–4220, 2019.

Z. Li and M. P. Wellman. A meta-game evaluation framework for deep multiagent reinforcement learning. In *33rd International Joint Conference on Artificial Intelligence (IJCAI)*, 2024.

R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *31st International Conference on Neural Information Processing Systems (NeurIPS)*, pages 6382–6393, 2017.

G. Papoudakis, F. Christianos, L. Schäfer, and S. V. Albrecht. Comparative evaluation of cooperative multi-agent deep reinforcement learning algorithms. *arXiv preprint arXiv:2006.07869*, 2020.

J. Perolat, R. Munos, J.-B. Lespiau, S. Omidshafiei, M. Rowland, P. Ortega, N. Burch, T. Anthony, D. Balduzzi, B. De Vylder, G. Piliouras, M. Lanctot, and K. Tuyls. From Poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *38th International Conference on Machine Learning (ICML)*, pages 8525–8535, 2021.

J. Perolat, B. D. Vylder, D. Hennes, E. Tarassov, F. Strub, V. de Boer, P. Muller, J. T. Connor, N. Burch, T. Anthony, S. McAleer, R. Elie, S. H. Cen, Z. Wang, A. Gruslys, A. Malysheva, M. Khan, S. Ozair, F. Timbers, T. Pohlen, T. Eccles, M. Rowland, M. Lanctot, J.-B. Lespiau, B. Piot, S. Omidshafiei, E. Lockhart, L. Sifre, N. Beauguerlange, R. Munos, D. Silver, S. Singh, D. Hassabis, and K. Tuyls. Mastering the game of Stratego with model-free multiagent reinforcement learning. *Science*, 378(6623):990–996, 2022.

T. Rashid, M. Samvelyan, C. S. De Witt, G. Farquhar, J. Foerster, and S. Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):7234–7284, 2020.

M. Schmid, M. Moravčík, N. Burch, R. Kadlec, J. Davidson, K. Waugh, N. Bard, F. Timbers, M. Lanctot, G. Z. Holland, et al. Student of games: A unified learning algorithm for both perfect and imperfect information games. *Science Advances*, 9(46):eadg3256:1–14, 2023.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419): 1140–1144, 2018.

A. Sinha and M. P. Wellman. Incentivizing collaboration in a competition.(2019). In *18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 13–17, 2019.

Y. Song, A. Wojcicki, T. Lukasiewicz, J. Wang, A. Aryan, Z. Xu, M. Xu, Z. Ding, and L. Wu. Arena: A general evaluation platform and building toolkit for multi-agent intelligence. In *34th AAAI Conference on Artificial Intelligence (AAAI)*, pages 7253–7260, 2020.

J. Stern. A chess formula is taking over the world. *The Atlantic*, Apr. 2024.

K. Tuyls, J. Perolat, M. Lanctot, E. Hughes, R. Everett, J. Z. Leibo, C. Szepesvári, and T. Graepel. Bounds and dynamics for empirical game-theoretic analysis. *Autonomous Agents and Multi-Agent Systems*, 34(7):1–30, 2020.

O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782): 350–354, 2019. doi: 10.1038/s41586-019-1724-z. URL `https://doi.org/10.1038/s41586-019-1724-z`.

Y. Vorobeychik and M. P. Wellman. Stochastic search methods for Nash equilibrium approximation in simulation-based games. In *7th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1055–1062, 2008.

M. P. Wellman. Putting the agent in agent-based modeling. *Autonomous Agents and Multi-Agent Systems*, 30:1175–1189, 2016.

B. Wiedenbeck, B.-A. Cassell, and M. P. Wellman. Bootstrap statistics for empirical games. In *13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 597–604, 2014.

P. R. Wurman, S. Barrett, K. Kawamoto, J. MacGlashan, K. Subramanian, T. J. Walsh, R. Capobianco, A. Devlic, F. Eckert, F. Fuchs, L. Gilpin, P. Khandelwal, V. Kompella, H. Lin, P. MacAlpine, D. Oller, T. Seno, C. Sherstan, M. D. Thomure, H. Aghabozorgi, L. Barrett, R. Douglas, D. Whitehead, P. Dürr, P. Stone, M. Spranger, and H. Kitano. Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature*, 602(7896):223–228, 2022.

C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *36th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 24611–24624, 2022.