Autonomously Reusing Knowledge in Multiagent Reinforcement Learning

Felipe Leno Da Silva[1], **Matthew E. Taylor**[2] and **Anna Helena Reali Costa**[1]
[1] University of São Paulo, São Paulo, Brazil
[2] Washington State University, Pullman, USA
f.leno@usp.br, taylorm@eecs.wsu.edu, anna.reali@usp.br

Abstract

Autonomous agents are increasingly required to solve complex tasks; hard-coding behaviors has become infeasible. Hence, agents must learn how to solve tasks via interactions with the environment. In many cases, knowledge reuse will be a core technology to keep training times reasonable, and for that, agents must be able to autonomously and consistently reuse knowledge from multiple sources, including both their own previous internal knowledge and from other agents. In this paper, we provide a literature review of methods for knowledge reuse in Multiagent Reinforcement Learning. We define an important challenge problem for the AI community, survey the existent methods, and discuss how they can all contribute to this challenging problem. Moreover, we highlight gaps in the current literature, motivating "low-hanging fruit" for those interested in the area. Our ambition is that this paper will encourage the community to work on this difficult and relevant research challenge.

1 Introduction

Knowledge reuse is of utmost importance when learning, both for humans and automated agents. Imagine a human child trying to learn how to play soccer. Given the task description, the child will unconsciously reuse previously acquired knowledge, such as walking while carrying a ball. If the child is not playing alone, other people (agents) might serve as additional knowledge sources to further accelerate learning. A more experienced teammate could demonstrate or explain moves that can be used to improve performance. Also, the behaviors and techniques of teammates or opponents may be observed and later imitated by the child. Additionally, if a coach is available, she can provide advice and rules for the child to quickly adapt to the team. Any professional soccer player will have taken advantage of all those knowledge sources multiple times, and a similar strategy is used by human beings to learn virtually any sequential decision-making task.

Humans successfully (and naturally) reuse previous internal knowledge, while simultaneously taking advantage of varied information received via implicit or explicit communication with other agents. However, the autonomous learning agents literature is primarily devoted to leveraging a few (often just one) of these possibly available knowledge sources, often in a predefined way. Autonomously identifying and using such sources on-the-fly imposes several challenges, including: computing task similarities, estimating individual agent performances for solving tasks, consistently combining multiple knowledge sources, and properly identifying the quality of knowledge received via communication.

This paper aims to encourage and support research to allow a learning agent to autonomously reuse knowledge in a multiagent system (MAS), identifying when other agents might provide demonstrations or advice and their willingness to help, as well as consistently combining and assimilating all the available information with the purpose of accelerating learning. We use reinforcement learning [Sutton and Barto, 1998] as the core learning technique for most of our discussion and references, but the ideas discussed may be applicable to many types of machine learning.

While reusing knowledge is not a novel idea, combining such a variety of information is an ambitious, challenging, and open research goal, with a potentially high payoff. One of the main challenges for multiagent *knowledge reuse* algorithms can be formulated as follows:

Create an agent able to autonomously identify and establish relationships with other agents that allow implicit or explicit knowledge sharing, and integrate the received information with its previous experience to improve learning.

This paper aims at surveying current lines of works towards solving this problem. Therefore, our contributions are to 1) introduce and motivate the general knowledge reuse challenge, 2) survey how the current literature from many subcommunities of AI contribute to addressing this challenge, and 3) identify areas where additional research is needed. The remainder of this paper is organized as follows. Section 2 presents the background on reinforcement learning and knowledge reuse. Section 3 formulates the learning problem. Section 4 presents the main categories of knowledge reuse techniques in the literature. The research questions are further elaborated in Section 5. Finally, suggestions for evaluation domains are presented in Section 6 and a long-term perspective for this challenge is discussed in Section 7.

2 Background

This section formulates sequential decision-making problems using a reinforcement learning (RL) framework and discusses how knowledge reuse can be successfully leveraged.

2.1 Markov Decision Processes and Multiagent Learning

Single-agent decision-making problems are usually formulated as *Markov Decision Processes* (MDPs). An MDP is composed of $\langle S, A, T, R \rangle$, i.e., a set of environment states S, a set of available actions A, a transition function T, and a reward function R. As the agent usually has no knowledge about R and T in learning problems, an MDP is solved through interacting with the environment and observing $\langle s, a, s', r \rangle$ tuples, where s is the state in which action a was applied, $s' = T(s, a)$ is the next state, and $r = R(s, a, s')$ is the received reward signal. Those *samples* of interactions are the feedback the agent has to solve the task. The goal of an RL agent is to induce a policy $\pi : S \rightarrow A$, that returns an action to be applied in each state. An optimal policy π^* always selects the action that maximizes the expected sum of rewards in a predefined horizon. Several algorithms to learn such a policy exist [Sutton and Barto, 1998].

MDPs are extended to MAS as *Stochastic Games* (SGs) [Busoniu *et al.*, 2008]. In an SG, S and A are defined as the Cartesian product of local states and actions of all agents, and each agent might have its own reward function. SGs are harder to solve, as now there is not a clear concept of optimal policy anymore. Agents might ignore the others in the environment and learn as in an MDP [Tan, 1993], try to learn an equilibrium joint policy [Hu *et al.*, 2015b], or try to learn a joint policy that maximizes a single common reward [Panait and Luke, 2005], depending on the problem to be solved. However, a common aspect of all those solutions is that the agent may require a significant amount of data to learn. Therefore, in order to be practically applied in complex domains, RL may need to be combined with additional speedup techniques, such as knowledge reuse.

2.2 Reuse of Knowledge

Reusing existing knowledge may accelerate learning, rendering complex tasks tractable. In order to solve an MDP or SG, the agent maps a knowledge space $\mathscr{K}$ to a policy $\pi \in \mathscr{H}$ [Lazaric, 2012], where $\mathscr{H}$ is the set of possible policies that can be chosen by the learning algorithm, $\mathscr{A} : \mathscr{K} \rightarrow \mathscr{H}$, where $\mathscr{K}$ is all the available knowledge the agent has to infer a policy and $\mathscr{A}$ is the algorithm. As explained in the previous section, when learning from scratch the set $\mathscr{K}$ contains only samples of interactions. However, when knowledge derived from previous tasks or additional sources are available, the agent can use the additional knowledge to solve the task: $\mathscr{A} : \mathscr{K}^{source} \cup \mathscr{K}^{agents} \cup \mathscr{K}^{target} \rightarrow \mathscr{H}$, where $\mathscr{K}^{source}$ is the knowledge from previous tasks, $\mathscr{K}^{agents}$ is the knowledge from other agents in the system[1], and $\mathscr{K}^{target}$ is the knowledge gathered from exploring the new task.

Algorithm 1 $Learn(Ag, \mathscr{K}^{source})$

```
for $\forall z \in REACHABLE(Ag)$ do
    $\mathscr{K}^{implicit} \leftarrow OBSERVE(z)$
    $\mathscr{K}^{explicit} \leftarrow ASK(z)$
$\mathscr{K}_C \leftarrow COMBINE(\mathscr{K}^{source}, \mathscr{K}^{implicit}, \mathscr{K}^{explicit})$
$REUSE(\mathscr{K}_C)$
```

However, deciding *when* and *what* to save in the knowledge bases, and *how* to reuse them to learn faster is a long-studied problem that has no single answer valid for all domains [Taylor and Stone, 2009]. Many papers have considered different ways to store and reuse useful information, following varied representations and assumptions, but they are usually reliant on a human designer to define how the knowledge should be reused. In the next section, we provide an integrated view on how the literature has been implementing knowledge reuse for MAS.

3 Problem Statement

Figure 1 illustrates the general case for knowledge reuse in MAS. Given a task to be solved, learning agents might reuse their own previous experience ($\mathscr{K}^{source}$) and also receive knowledge from other agents ($\mathscr{K}^{agents}$), which may be also learning. Those knowledge sources might be combined with experiences gathered from exploring the new task ($\mathscr{K}^{target}$), and after solving the task the agent can update its knowledge base for reusing it.

Algorithm 1 presents an algorithmic (and simplified) view of how the agent should learn. Let Ag be the set of agents in the MAS and $\mathscr{K}^{source}$ be the knowledge available from previously solved tasks. For each reachable[2] agent z, our agent will observe z and evaluate if knowledge can be implicitly extracted. For example, even if the agent cannot communicate with z, it might be possible to extract knowledge by observing how it acts, which we define as *implicit* knowledge reuse. $\mathscr{K}^{implicit}$ is the knowledge extracted from such an observation. Then, our agent communicates with z and checks if z has useful knowledge to send via explicit knowledge communication, such as advice. We label any received knowledge $\mathscr{K}^{explicit}$. Finally, the agent needs to combine all the available knowledge and reuse it to improve its policy. It is noteworthy that this process is not only executed a single time, but rather periodically repeated until the agent needs no addition of knowledge for solving the task.

In practice, simultaneously leveraging knowledge from both previous experience and other agents is not a trivial task, as the generated data is usually heterogeneous and it is not easy to cope with potential conflicts that may arise between $\mathscr{K}^{source}$, $\mathscr{K}^{agents}$, and $\mathscr{K}^{target}$. Partially for that reason, most of the literature focuses on one of the knowledge sources alone for reusing knowledge. In the next section we discuss the most relevant categories of approaches in the literature.

[1] Other agents, including humans, could be actively participating in the system in a SG. Agents that are watching the learning agent and providing advice (but not interacting directly with the environment) apply in both MDPs and SGs.

[2] For some domains, agents may be unreachable in some situations (e.g., they are too far to communicate and/or observe).

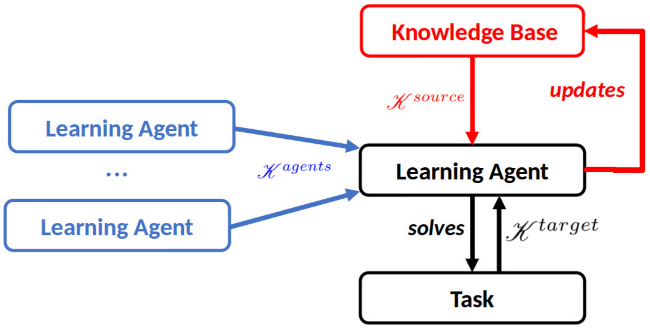

Figure 1: Illustration of the general knowledge reuse problem for MAS. A learning agent might receive knowledge from communications with other agents ($\mathscr{K}^{agents}$) and/or reuse its previous knowledge ($\mathscr{K}^{source}$) for combining it with exploration in the target task ($\mathscr{K}^{target}$).

4 Knowledge Reuse Strategies

In this section we introduce the main categories of knowledge reuse in the existing literature and give references to representative works.

4.1 Knowledge from Previous Tasks

If an agent has to solve more than one task in its lifetime, knowledge from previously solved tasks (i.e., *source tasks*) may be reused to solve a new task (i.e., a *target task*). Previous works have used various methods to transfer knowledge between tasks. Samples [Lazaric *et al.*, 2008; Tan, 1993; Taylor *et al.*, 2008], partial policies [Topin *et al.*, 2015], abstracted knowledge [Koga *et al.*, 2015], value functions [Taylor *et al.*, 2014a; 2007], and heuristics or biases [Bianchi *et al.*, 2015; Boutsioukis *et al.*, 2011] might all be used. However, none of those methods dominates the others in all situations and automatically selecting which method will perform the best is a difficult open problem. When learning on adversarial tasks, the knowledge gained by playing against an opponent can be adapted to other ones [Hernandez-Leal and Kaisers, 2017; Kelly and Heywood, 2015], but for that the agent must be able to estimate the similarity between opponents.

4.2 Learning from Demonstration

A *teacher* can provide demonstrations to a less experienced agent (i.e., a *student*).[3] Demonstrations may be in the form of teleoperation or instructions from the teacher for a certain period of time, usually resulting in a set of interactions with the environment following the teacher's policy. The student can then incorporate this set by building a predictive model of the mentor's policy [Ude *et al.*, 2004; Chernova and Veloso, 2009] or by estimating reward or transition models to accelerate learning [Thomaz and Breazeal, 2006; Abbeel *et al.*, 2007]. Those mentoring relations are usually fixed only at the beginning of learning, but ideally they should be initiated at strategic times, such as when the student has a high uncertainty on what to do in the current state, so as to avoid wasting the teacher's attention in situations for which the student has a good policy [Chernova and Veloso, 2009]. A flexible method should also consider that the teacher's policy might be suboptimal and enable student exploration.

4.3 Imitation Learning

When no direct communication between agents is available, either because no common message passing protocol is known or because one agent is not willing to share its knowledge with another, behaviors can still be learned through observation of another agent's actuation. When solving a task, if observation is possible, an experienced agent can implicitly communicate information about its policy that may be used by an observer, even if the observer cannot precisely identify which action was taken [Price and Boutilier, 2003; Shon *et al.*, 2007]. The experienced agent may be acting with the intent of teaching a less experienced agent, or it may be following a policy without knowing (or caring) that another agent is watching. It is not easy to identify when it is worth trying to imitate an agent, as the observer may not be able to know if the other agent has the same reward function or even if its policy is good without trying to imitate it in the environment.

4.4 Advising

Action advising is one of the most flexible knowledge reuse procedures, as agents need only a common understanding of the actions and current state. The main idea is to have one (or multiple [Taylor *et al.*, 2014b]) *teacher* agent, which provides action suggestions to a *student*. While most works focus only on either how to best profit from human advice [Maclin and Shavlik, 1996; Amir *et al.*, 2016] or advice between artificial agents [Taylor *et al.*, 2014b; Silva *et al.*, 2017], a comprehensive advising framework would treat any agent equally, regardless of its "implementation," as long as some protocol is followed for providing the suggestions. As communication is usually limited or costly, identifying precisely when to provide action suggestions is critical. While some works leave the burden of initiating advising relations either on the teacher or the student, ideally, advice would only be provided when both agents agree it is important for the current situation [Amir *et al.*, 2016; Silva *et al.*, 2017].

4.5 Curriculum Learning

Curriculum learning is typically used to decompose a hard (target) task into several easier (source) tasks, and learning them in a convenient sequence [Narvekar *et al.*, 2016]. When appropriate, task decomposition allows the learning agent to use knowledge from the source tasks (which are quickly solvable), which can be reused to 1) accelerate learning in the target task, such that it can be learned faster with the curriculum than without; or 2) accelerate learning on the sequence of tasks, such that learning the entire curriculum is faster than directly learning on the target task without the curriculum. However, defining an appropriate sequence (i.e., the curriculum) is hard both for humans [Peng *et al.*, 2016b] and autonomous agents [Silva and Costa, 2018; Narvekar *et al.*, 2017]. A current challenge in this subarea

[3] The teacher referenced in Sections 4.2–4.4 can be a human or an artificial agent.

is designing agents capable of building curricula for other agents [Sukhbaatar *et al.*, 2018; Matiisen *et al.*, 2017], which should ideally be built taking into account the individual capabilities of the learning agent [Narvekar *et al.*, 2017].

4.6 Additional Related Areas

Slightly different settings of the knowledge reuse problem are tackled under different names. Transfer learning [Taylor and Stone, 2009] is the knowledge reuse across pairs of predefined tasks. Multi-task learning consists of starting the learning process with a knowledge base of several already solved tasks, which can be used in the new task. Lifelong learning [Isele *et al.*, 2016] consists of reusing and adapting knowledge across tasks that are presented to the learning agent during its life-span. Zero-Shot Learning (under the RL perspective) [Isele *et al.*, 2016; Silva and Costa, 2017] aims at reusing previous knowledge before training in the new task. All those subareas are closely related and can be fit in the category of *Knowledge from Previous Tasks* (Section 4.1). This challenge is also related to *Ad Hoc Teamwork* [Stone *et al.*, 2010], which studies how to autonomously coordinate with previously unknown teammates because *ad hoc* knowledge-sharing relations will be needed.

5 Open Problems and Partial Solutions

In order to tackle the learning problem presented in Section 3, several difficult open research questions must be answered. We now list a selection of promising lines of work.

- **Detection of Knowledge Reuse Opportunities**: Identifying the right moment to reuse knowledge likely has a major influence on the success of a technique. Even if an agent is willing to cooperate and agrees to provide advice, for example, it will only make sense in cases where the teacher has more knowledge than the student in the current situation. Similarly, imitating an agent that has a worse performance will probably hamper the learning process instead of accelerating it. However, evaluating the learning performance online is sometimes hard, even in cooperative tasks (because it may be hard to identify which agent is responsible for the current reward). While some existing work aims at activating the knowledge reuse technique only when the agent is uncertain on its policy [Silva *et al.*, 2017; Peng *et al.*, 2016a; Amir *et al.*, 2016], most of the literature assumes that the available knowledge is useful and provided by an expert agent [Taylor and Stone, 2009], which is not necessarily true. The agent must be able to identify if the knowledge source is expected to improve its own performance and to contact (or observe) other agents to establish an appropriate knowledge reuse relationship.
- **Estimation of Task Similarities**: In order to reuse the previous internal knowledge, the agent needs to find the previously solved tasks that are most similar to the new task, avoiding *negative transfer* [Taylor and Stone, 2009]. Selecting only appropriate knowledge is especially important for our challenge, as inconsistent knowledge may be hard to overwrite. Some papers contributed techniques to autonomously compute task similarity [Isele *et al.*, 2016] but their applicability is usually limited and they may be hard to combine with varied knowledge reuse techniques.
- **Consistent Combination of Knowledge Sources**: Simultaneously applying multiple knowledge reuse techniques creates a new problem to be solved. If the agent decides to reuse a policy from a previous task and other agents are willing to provide demonstrations and advice at the same time, should the agent use all the knowledge sources simultaneously or select only a subset of the possible methods? How should the agent combine and assimilate this information? Learning through combinations of such heterogeneous data is a novel research problem.
- **Reputation and Knowledge Quality Detection**: Identifying the correct timing and the correct partner for receiving knowledge may not be enough in self-interested MAS. Virtually all the approaches in the literature assume that all agents in knowledge reuse interactions are benevolent, i.e., they will never communicate suboptimal information on purpose. Often, not even the possibility of having *teachers* with imperfect knowledge is considered, simply taking the communicated knowledge for optimal, as in [Chernova and Veloso, 2009]. However, for most of real-world problems, benevolence cannot be assumed. Malicious communications could possibly transmit incorrect or inconsistent knowledge with the intent of harming the agent or reducing its performance. Hence, the agents might be required to negotiate knowledge [Shon *et al.*, 2007] and implement a trust mechanism for avoiding catastrophic damages in case of receiving malicious communications.
- **Performance Metric**: Evaluating even a single knowledge reuse method is not a simple task. Multiple performance metrics exist, such as the *asymptotic performance*, *time to threshold*, and *jumpstart* [Taylor and Stone, 2009]. Each of them evaluates an aspect of the impact of introducing knowledge in the learning process. However, other metrics are relevant for MAS depending on the setting, such as resources expended for negotiating, resources needed to process knowledge, or the amount of human attention required. Partially because of the lack of unifying metrics, evaluations are often subjective, which hinders the comparison of methods. To have a clear picture of the knowledge reuse method, new metrics especially tailored to the MAS setting are needed.
- **Theoretical Analysis of Knowledge Reuse**: Although some theoretically-inspired works do exist [Zhan *et al.*, 2016; Isele *et al.*, 2016], knowledge reuse approaches for RL have been predominantly empirical. Developing theoretical approaches with known bounds on regret and resources used would be a major development for the area as a whole.
- **Team-Oriented Communication**: Most of the literature focuses on communicating knowledge in tasks where the local state of agents is only affected by their

own actions [Price and Boutilier, 2003; Chernova and Veloso, 2009]. However, when agents must coordinate, the communicated knowledge should be adapted to the role that the receiving agent is expected to play, instead of following the policy of the agent communicating it. However, how to atomatically adapt internal knowledge for encouraging coordination is still an open problem.

- **Knowledge Exchange Protocols**: The protocols for negotiating and transferring knowledge are implicitly hardcoded in the codification of the agent in most of the literature [Taylor and Stone, 2009]. However, knowledge reuse for autonomous agents in the real world requires the creation of explicit protocols, in order to enable a new agent to quickly and easily join the MAS, participating in and understanding communications.

6 Domain Suggestions

There are many possible domains in which the ideas and challenges above can be tested. The primary requirement is that multiple types of agents can learn to perform a task, but *desiderata* include the following: 1) humans can learn to perform the task; 2) an interface is available for human understanding and interaction; 3) if the task is real time, the speed of the task is variable to ensure human and agent playability; 4) the ability for agents to see each other and communicate; 5) the existence of multiple scenarios, with varying degrees of similarity and difficulty; and 6) multiple types of high-performing policies in a given task and/or non-obvious optimal policies.

We suggest several domains that can meet all of the above criteria. A simple, quickly deployable gridworld domain, a more complex domain of simulated robot soccer, and the real application of electric vehicle coordination.

6.1 Multiagent Gridworld

The *Gridworld* domain and its variations have been extensively used for evaluations of multiagent learning algorithms [Price and Boutilier, 2003; Hu *et al.*, 2015a; Hernandez-Leal and Kaisers, 2017]. This domain is easy to implement, customize, and to interpret results, and therefore has been a standard "first trial" domain for MAS. Possible evaluation settings include:

- **Competitive Resource Gathering**: Limited resources are spread in the environment, together with two opposing teams of agents. The learning agent does not have initial knowledge about the policies of any of the other agents in the environment but can try to negotiate knowledge with its teammates or to imitate successful strategies used by any of the teams. For building more complex versions of the task, different kinds and numbers of resources might be spread in the environment.

- **Progressive Predator-Prey**: Predators aim to capture prey. A team of learning predators has the goal of capturing groups of prey that become progressively smarter. Knowledge gained from previously successful strategies could be reused, and predators could choose to share knowledge among themselves.

6.2 Simulated Robot Soccer

The RoboCup task [Kitano *et al.*, 1997] is an ideal real-time testbed for such knowledge reuse techniques. Strategies and moves might be imitated from teammates or opponents, communication can be established to transfer knowledge, teleoperation is possible, and knowledge can be reused from previous games [Bianchi *et al.*, 2009]. As deploying the full RoboCup simulation may be hard, simplifications of the robot soccer task such as the *Keepaway* [Stone *et al.*, 2005] and *Half Field Offense* [Hausknecht *et al.*, 2016] environments might also be used.

6.3 Electric Vehicle Coordination

The *Electric Vehicle Charging* coordination problem [Taylor *et al.*, 2014a; Silva and Costa, 2015] is one of the real-world applications in which multiagent RL was successfully employed. This problem consists of a neighborhood connected to a single transformer that provides energy for all houses. Each of the houses has an *Electric Vehicle* (EV), which must be recharged for daily use. However, as most of people have similar work shifts, if they do not refrain from recharging their EVs right after arriving home, the transformer limit load is likely to be exceeded in peak hours, causing undesired infrastructural damages or costs. Therefore, the learning problem consists of implementing each of the EVs as an autonomous RL agent, which aims to recharge a car while coordinating with other EVs to avoid overloads in the network. The agents must also take into account the energy demand of the houses, which cannot be modified.

Initial works [Taylor *et al.*, 2014a] succeeded in transferring knowledge between agents, but many knowledge reuse opportunities remain unexplored, such as reusing knowledge of EVs in the neighborhood for deriving an initial policy for newcomers.

7 Long-Term Perspectives and Conclusion

This paper discusses and surveys a challenging research problem: Development of methods for *autonomously* identifying and taking advantage of *knowledge reuse* opportunities, both between *tasks* and *agents*. Although there are currently many techniques to reuse knowledge, they typically require a human to hard-code which knowledge source is available and how to use it. Developing protocols for knowledge negotiation and transfer will be critical to achieve the proposed goals. Ideally, this common protocol would be human-interpretable, allowing human-machine collaboration through knowledge-exchange mechanisms. Security is also a key concern for scaling learning MAS but has been neglected by most of the knowledge reuse literature. In the real world, agents must not assume the benevolence of the others, thus mechanisms for assessing the quality of received knowledge before assimilating it are indispensable for real applications, avoiding undesirable effects caused by malicious communications.

Adapting knowledge for the specificity of each agent is also a major challenge for the area. The current literature does not provide techniques for adapting the knowledge built by an agent to another with different sensors and/or actuators, but the community is slowly moving towards this problem.

In the short-term, we hope that this survey will direct community members to make use of all the knowledge available in a MAS, and not just portions of it. In the long-term, we expect that agents will be able to autonomously develop protocols to knowledge exchange and negotiation in a way that it will be still explainable to humans, creating a new generation of intelligent agents.

Acknowledgments

We gratefully acknowledge financial support from CNPq, grants 311608/2014-0 and 425860/2016-7 and São Paulo Research Foundation (FAPESP), grants 2015/16310-4, 2016/21047-3, and 2018/00344-5.

References

[Abbeel *et al.*, 2007] Pieter Abbeel, Adam Coates, Morgan Quigley, and Andrew Y. Ng. An Application of Reinforcement Learning to Aerobatic Helicopter Flight. In *Advances in Neural Information Processing Systems*, pages 1–8, 2007.

[Amir *et al.*, 2016] Ofra Amir, Ece Kamar, Andrey Kolobov, and Barbara Grosz. Interactive Teaching Strategies for Agent Training. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 804–811, 2016.

[Bianchi *et al.*, 2009] Reinaldo Bianchi, Raquel Ros, and Ramon Lopez de Mantaras. Improving reinforcement learning by using case based heuristics. *Case-Based Reasoning Research and Development*, pages 75–89, 2009.

[Bianchi *et al.*, 2015] Reinaldo Bianchi, Luiz A. Celiberto Jr., Paulo E. Santos, Jackson P. Matsuura, and Ramon Lopez de Mantaras. Transferring Knowledge as Heuristics in Reinforcement Learning: A Case-Based Approach. *Artificial Intelligence*, 226:102 – 121, 2015.

[Boutsioukis *et al.*, 2011] Georgios Boutsioukis, Ioannis Partalas, and Ioannis Vlahavas. Transfer Learning in Multi-agent Reinforcement Learning Domains. In *European Workshop on Reinforcement Learning*, 2011.

[Busoniu *et al.*, 2008] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(2):156–172, 2008.

[Chernova and Veloso, 2009] Sonia Chernova and Manuela Veloso. Interactive Policy Learning through Confidence-Based Autonomy. *Journal of Artificial Intelligence Research (JAIR)*, 34(1):1, 2009.

[Hausknecht *et al.*, 2016] Matthew Hausknecht, Prannoy Mupparaju, Sandeep Subramanian, Shivaram Kalyanakrishnan, and Peter Stone. Half Field Offense: An Environment for Multiagent Learning and Ad Hoc Teamwork. In *AAMAS Adaptive Learning Agents (ALA) Workshop*, 2016.

[Hernandez-Leal and Kaisers, 2017] Pablo Hernandez-Leal and Michael Kaisers. Towards a Fast Detection of Opponents in Repeated Stochastic Games. In *Workshop on Transfer in Reinforcement Learning (TiRL)*, 2017.

[Hu *et al.*, 2015a] Yujing Hu, Yang Gao, and Bo An. Learning in Multi-agent Systems with Sparse Interactions by Knowledge Transfer and Game Abstraction. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 753–761, 2015.

[Hu *et al.*, 2015b] Yujing Hu, Yang Gao, and Bo An. Multiagent Reinforcement Learning with Unshared Value Functions. *IEEE Transactions on Cybernetics*, 45(4):647–662, 2015.

[Isele *et al.*, 2016] David Isele, Mohammad Rostami, and Eric Eaton. Using Task Features for Zero-Shot Knowledge Transfer in Lifelong Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1620–1626, 2016.

[Kelly and Heywood, 2015] Stephen Kelly and Malcolm I Heywood. Knowledge Transfer from Keepaway Soccer to Half-Field Offense through Program Symbiosis: Building Simple Programs for a Complex Task. In *Conference on Genetic and Evolutionary Computation (GECCO)*, pages 1143–1150, 2015.

[Kitano *et al.*, 1997] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, Eiichi Osawa, and Hitoshi Matsubara. Robocup: A challenge problem for AI. *AI magazine*, 18(1):73, 1997.

[Koga *et al.*, 2015] Marcelo Li Koga, Valdinei Freire da Silva, and Anna Helena Reali Costa. Stochastic Abstract Policies: Generalizing Knowledge to Improve Reinforcement Learning. *IEEE Transactions on Cybernetics*, 45(1):77–88, 2015.

[Lazaric *et al.*, 2008] Alessandro Lazaric, Marcello Restelli, and Andrea Bonarini. Transfer of Samples in Batch Reinforcement Learning. In *International Conference on Machine Learning (ICML)*, pages 544–551, 2008.

[Lazaric, 2012] Alessandro Lazaric. *Transfer in Reinforcement Learning: A Framework and a Survey*, pages 143–173. Springer Berlin Heidelberg, 2012.

[Maclin and Shavlik, 1996] Richard Maclin and Jude W. Shavlik. Creating Advice-Taking Reinforcement Learners. In *Machine Learning*, pages 251–281, 1996.

[Matiisen *et al.*, 2017] Tambet Matiisen, Avital Oliver, Taco Cohen, and John Schulman. Teacher-Student Curriculum Learning. *arXiv:1707.00183*, 2017.

[Narvekar *et al.*, 2016] Sanmit Narvekar, Jivko Sinapov, Matteo Leonetti, and Peter Stone. Source Task Creation for Curriculum Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 566–574, 2016.

[Narvekar *et al.*, 2017] Sanmit Narvekar, Jivko Sinapov, and Peter Stone. Autonomous Task Sequencing for Customized Curriculum Design in Reinforcement Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2536–2542, 2017.

[Panait and Luke, 2005] Liviu Panait and Sean Luke. Cooperative Multi-Agent Learning: The State of the Art. *Au-*

tonomous Agents and Multiagent Systems, 11(3):387–434, 2005.

[Peng *et al.*, 2016a] Bei Peng, James MacGlashan, Robert Loftin, Michael L. Littman, David L. Roberts, and Matthew E. Taylor. A Need for Speed: Adapting Agent Action Speed to Improve Task Learning from Non-Expert Humans. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 957–965, 2016.

[Peng *et al.*, 2016b] Bei Peng, James MacGlashan, Robert Loftin, Michael L. Littman, David L. Roberts, and Matthew E. Taylor. An Empirical Study of Non-expert Curriculum Design for Machine Learners. In *IJCAI Interactive Machine Learning Workshop*, 2016.

[Price and Boutilier, 2003] Bob Price and Craig Boutilier. Accelerating Reinforcement Learning through Implicit Imitation. *Journal of Artificial Intelligence Research (JAIR)*, 19:569–629, 2003.

[Shon *et al.*, 2007] Aaron P. Shon, Deepak Verma, and Rajesh P. N. Rao. Active Imitation Learning. In *AAAI Conference on Artificial Intelligence*, pages 756–762, 2007.

[Silva and Costa, 2015] Felipe Leno Da Silva and Anna Helena Reali Costa. Multi-Objective Reinforcement Learning through Reward Weighting. In *Workshop on Synergies Between Multiagent Systems, Machine Learning and Complex Systems (TRI) at IJCAI*, volume 1, pages 25 – 36, 2015.

[Silva and Costa, 2017] Felipe Leno Da Silva and Anna Helena Reali Costa. Towards Zero-Shot Autonomous Inter-Task Mapping through Object-Oriented Task Description. In *Workshop on Transfer in Reinforcement Learning (TiRL)*, 2017.

[Silva and Costa, 2018] Felipe Leno Da Silva and Anna Helena Reali Costa. Object-Oriented Curriculum Generation for Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018.

[Silva *et al.*, 2017] Felipe Leno Da Silva, Ruben Glatt, and Anna Helena Reali Costa. Simultaneously Learning and Advising in Multiagent Reinforcement Learning. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1100–1108, 2017.

[Stone *et al.*, 2005] Peter Stone, Richard S. Sutton, and Gregory Kuhlmann. Reinforcement Learning for RoboCup-Soccer Keepaway. *Adaptive Behavior*, 13(3):165–188, 2005.

[Stone *et al.*, 2010] Peter Stone, Gal A. Kaminka, Sarit Kraus, and Jeffrey S. Rosenschein. Ad Hoc Autonomous Agent Teams: Collaboration without Pre-Coordination. In *AAAI Conference on Artificial Intelligence*, pages 1504–1509, 2010.

[Sukhbaatar *et al.*, 2018] Sainbayar Sukhbaatar, Ilya Kostrikov, Arthur Szlam, and Rob Fergus. Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play. In *International Conference on Learning Representations (ICLR)*, 2018.

[Sutton and Barto, 1998] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[Tan, 1993] Ming Tan. Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents. In *10th International Conference on Machine Learning (ICML)*, pages 330–337, 1993.

[Taylor and Stone, 2009] Matthew E. Taylor and Peter Stone. Transfer Learning for Reinforcement Learning Domains: A Survey. *Journal of Machine Learning Research*, 10:1633–1685, 2009.

[Taylor *et al.*, 2007] Matthew E. Taylor, Peter Stone, and Yaxin Liu. Transfer Learning via Inter-Task Mappings for Temporal Difference Learning. *Journal of Machine Learning Research*, 8(1):2125–2167, 2007.

[Taylor *et al.*, 2008] Matthew E. Taylor, Nicholas K. Jong, and Peter Stone. Transferring Instances for Model-Based Reinforcement Learning. In *Machine Learning and Knowledge Discovery in Databases*, volume 5212 of *Lecture Notes in Artificial Intelligence*, pages 488–505, 2008.

[Taylor *et al.*, 2014a] Adam Taylor, Ivana Dusparic, Edgar Galvan-Lopez, Siobhan Clarke, and Vinny Cahill. Accelerating Learning in Multi-Objective Systems through Transfer Learning. In *International Joint Conference on Neural Networks (IJCNN)*, pages 2298–2305, July 2014.

[Taylor *et al.*, 2014b] Matthew E. Taylor, Nicholas Carboni, Anestis Fachantidis, Ioannis P. Vlahavas, and Lisa Torrey. Reinforcement Learning Agents Providing Advice in Complex Video Games. *Connection Science*, 26(1):45–63, 2014.

[Thomaz and Breazeal, 2006] Andrea Lockerd Thomaz and Cynthia Breazeal. Reinforcement Learning with Human Teachers: Evidence of Feedback and Guidance with Implications for Learning Performance. In *AAAI Conference on Artificial Intelligence*, 2006.

[Topin *et al.*, 2015] Nicholay Topin, Nicholas Haltmeyer, Shawn Squire, John Winder, Marie desJardins, and James MacGlashan. Portable Option Discovery for Automated Learning Transfer in Object-Oriented Markov Decision Processes. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3856–3864, 2015.

[Ude *et al.*, 2004] Aleš Ude, Christopher G. Atkeson, and Marcia Riley. Programming Full-body Movements for Humanoid Robots by Observation. *Robotics and Autonomous Systems*, 47(2):93 – 108, 2004.

[Zhan *et al.*, 2016] Yusen Zhan, Haitham Bou-Ammar, and Matthew E. Taylor. Theoretically-Grounded Policy Advice from Multiple Teachers in Reinforcement Learning Settings with Applications to Negative Transfer. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2315–2321, 2016.