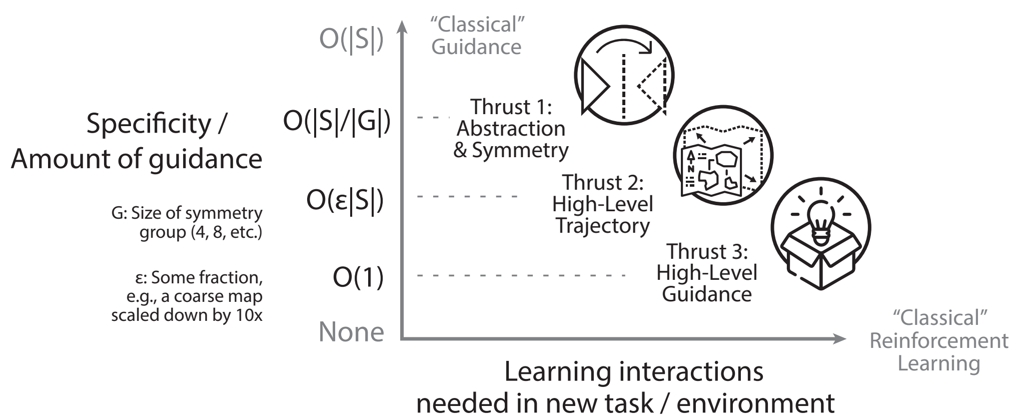
Research Statement

Lawson L.S. Wong, Khoury College of Computer Sciences, Northeastern University

Robots and autonomous systems need to learn about their environment and adapt their behavior quickly, in order to operate effectively in unstructured and diverse environments such as homes and offices. Reinforcement learning (RL) is a very general and promising technique for enabling agents to learn and adapt. In principle, RL algorithms can learn to act optimally on any task and environment, without needing any prior information. However, this comes at a prohibitive cost – fundamentally, they must try all actions in all states to learn how the universe works and what behavior is best. My thesis is that RL is difficult because it is too general; we need to, and often can, provide RL a helping hand. Rather than solving the most general RL problem, we may be able to make RL much more tractable by using a modicum of task-relevant high-level information.

In practice, to make real RL systems learn quickly, expert users often rely on reward shaping and imitation learning. We consider these approaches to be providing *low-level guidance*, because they require labeling what is good (or bad) directly in the agent's (low-level) state and action spaces respectively. These forms of guidance are difficult to specify and get right, even for experts. In contrast, humans can use *high-level guidance* to help them solve new tasks and avoid extensive trial-and-error learning. For example, we can recognize situations equivalent to ones we have seen before; we can navigate in novel environments using coarse 2-D maps; we can cook new dishes by following recipes; we can follow partial advice such as "avoid driving too close to the cliffside". What these examples have in common, and which is different from the low-level guidance mentioned above, is that the extra input does not have a direct correspondence to the agent's state-action space, and may not even be immediately relevant to solving the task. These forms of high-level guidance still clearly contain useful information, but RL algorithms cannot effectively use them.

At Northeastern, I lead the Generalizable Robotics and Artificial Intelligence Laboratory (GRAIL), where we make progress on this agenda via multiple thrusts. High-level guidance comes in many forms and lie on a spectrum illustrated below. Overall, there is a trade-off between the amount of guidance specified and the sample complexity of learning a new task / in a new environment. In the first thrust, we use symmetry to identify equivalent states/actions, cutting down the amount of guidance by the order of the relevant symmetry group. The second thrust focuses on approaches for robustly following a single high-level trajectory, such as a path on a coarse 2-D map; the amount of guidance is further reduced to a small fraction of the state space, since only a single rough trajectory is provided. In the future, our third thrust considers even more flexible forms of guidance, such as a constant amount of natural language "hints". In this case, we expect that the agent will still need to do most of the learning from interacting with the environment, but if the guidance is used to explore effectively, we may already reduce training time by an order of magnitude or more.



Abstraction and symmetry in reinforcement learning

At its core, the question of generalization relates to finding a good representation or *abstraction*: What are the things that matter, and more importantly, what details are irrelevant? With sufficient detail, every environment and task is unique. Generalization is only achievable when the similarities between situations are exposed, and irrelevances hidden, at appropriate levels of abstraction.

To study this problem formally, we consider the problem of state abstraction in Markov decision processes (MDPs) and RL, where the set of possible states (and actions) is typically too large to enumerate and learn individually from; they must be aggregated in some way. By adapting information-theoretic bounds and algorithms to the MDP setting, we proposed new algorithms and bounds for state abstraction [1]. To scale up our state-abstraction approach, we developed a new variational information-bottleneck method that learns a low-dimensional continuous encoding of image-based states, similar to variational autoencoders. This encoding is further used to extract a small discrete abstract MDP that is an approximate bisimulation of the original MDP [2]. This allows us to apply dynamic programming algorithms on the abstract MDP to obtain policies for new tasks, which can then be efficiently translated into the original MDP via the learned bisimulation.

Another way to generalize between different problems is by recognizing the symmetry between them. For example, if we know that we should swerve left to avoid collisions in a certain configuration of cars, when presented with a mirror image of the configuration, we should automatically infer to swerve right. Generally, incorporating the correct symmetry (equivariances) can lead to significant gains in learning efficiency, achieving a speedup roughly on the order of the size of the symmetry group G, and possibly even exceeding the performance of non-equivariant methods.

In RL, one way to incorporate symmetry is by constraining the value/policy functions using appropriate equivariant neural network architectures. In contrast to these model-free methods, we have explored a model-based framework for integrating symmetry into differentiable planning algorithms. We started with Value Iteration Networks (VIN), a differentiable form of value iteration. By constraining each step (layer) of VINs to be equivariant to the domain symmetry, we thus enforce that all intermediate computations of planning, and thus the entire planning algorithm, are constrained by symmetry [3]. To keep the deep planning network stable to train, we further applied implicit differentiation techniques [4]. These advances lead to significantly faster model-learning and higher success rates on grid-based planning domains, including domains with visual input. We have also extended this work to operate on graphs instead of grids, with a novel equivariant lifting method, such that it is more applicable to robot/embodied visual navigation [5]. Finally, we have investigated how to create equivariant sampling-based differentiable planning algorithms [6].

Our study of equivariant models for decision making has yielded other fruitful findings. Many RL problems have factorial structure, such as objects. When factored in this way, our learned world models should be able to recombine previously learned factors to solve novel tasks involving new combinations of objects, without any further training required (e.g., build a new configuration of blocks) [7, 8]. We used permutation equivariance to derive theoretical guarantees of this form of combinatorial generalization [9]. We have also addressed a common perceived limitation of equivariant models, specifically when the inductive bias of symmetry does not hold. Surprisingly, this is not as harmful as expected – our equivariant methods can work well even when the assumed symmetry did not hold completely in the domain. To investigate this further, we proposed a notion of extrinsic symmetry [10] (where the wrong inductive bias is irrelevant), to distinguish from incorrect symmetry (which we prove can be harmful with additional assumptions). Further afield, we have proposed gauge-equivariant models for solving partial differential equations on meshes [11].

This work is supported by NSF Award #2107256 (RI: Medium: Learning MDP abstractions for autonomous systems using variational methods and symmetry groups, $1.2M).

Following a single high-level trajectory

Even with the right representations, finding a good solution can be still be difficult because of large state/action spaces, long horizons, and sparse rewards. Fundamentally, the RL approach of trial-and-error learning is challenging under these circumstances. When a tourist tries to find their way in a new city, it would be futile to do so by trial and error; instead, they consult a map, plan a path, and follow it to their destination. Similarly, we might watch a video to learn a new kitchen/laboratory skill, or follow steps in a manual to assemble furniture. The common theme in all these examples is that a *single high-level trajectory* is provided, outlining the solution but not the fine details. The central question of this thrust is: If we provide a *rough* solution of a problem to an agent, can the agent learn to follow the solution effectively and robustly?

This question resembles learning from demonstration and one-shot imitation learning, which are popular alternatives to RL. However, demonstrations are typically provided directly in the agent's low-level action space. Additionally, imitation learning approaches suffer from covariate shift; once the agent deviates from the demonstrated region, the agent behaves erratically and cannot recover.

In contrast, we propose a general approach that only requires a single high-level trajectory, provided in a different state space and/or with different agents/objects, and tries to follow it as strictly as possible. Specifically, we have been studying *abstract navigation* problems, where an agent is situated within an environment whose layout has never been seen before, and the agent is expected to navigate to a goal without first training on or even exploring this domain. This task may appear impossible without further guidance, but we provide the agent with a coarse 2-D map of the environment, as well as indicators of its start and goal locations.

The key insight is recognizing that the solution is essentially contained on the coarse map; all that is missing is how to translate between the real world and the abstract / high-level space. Going from the real world to the coarse map, we treat this as a localization problem, where we use egocentric visual observations to estimate the agent's location on the coarse map [12]. This involves learning an observation model that connects the real and abstract spaces, and maintaining estimates with a Bayesian filter. Once the agent is sufficiently well-localized on the coarse map, a high-level trajectory of subgoals can be found using graph search. Each subgoal is then given to a low-level controller that is trained (with goal-conditioned RL) to robustly reach nearby landmarks in the agent's visually rich and egocentric perspective [13]. We have validated this pipeline both in photorealistic simulators and on a physical mobile robot. Besides this structured approach, we have also proposed a more end-to-end approach that treats the abstract map as an unstructured image, and learns to plan using the supplied map [14]. The correspondence between the abstract map and the agent is implicitly learned via a map-conditioned transition function.

Beyond navigation, we have studied a related one-shot imitation learning problem in robot manipulation. In this case, while the single demonstration is not necessarily more abstract, it is still high-level in the sense that test situations will involve objects with different shapes and poses. We proposed an approach that extracts interaction keypoints from the demonstration trajectory, and warps them to fit model parameters inferred from observations of the novel object/pose [15].

Additionally, to mitigate effects of covariate shift, we developed an approach that encourages the agent to return to the single demonstration whenever it deviates from it [16]. We learned a reverse-time model that generates trajectories heading away from the demonstration, which, when reversed, yields segments that converge toward the demonstration. Augmenting the training data in this way produces "funnels" that expand the basin of attraction of the single demonstration.

This work is supported by NSF Award #2142519 (Collaborative research: Design, flight control, and autonomous navigation of bioinspired morphing micro aerial vehicles for operation in confined spaces, $652K), a Northeastern TIER 1 seed grant ($50K), and a Khoury seed grant ($50K).

Robotics

My group is ultimately interested in robot generalization, hence we maintain many active projects in robotics to validate and further inspire our approaches. We have developed approaches for abstract robot navigation and mapping as described in the previous section [12, 13, 14], integrating with symmetry as well [5]. In addition to the interaction warping work above [15], our contributions to robot manipulation include a hierarchical imitation learning approach for bimanual robot manipulation [17], and the use of action priors to speed up RL exploration in tabletop manipulation tasks where action spaces are large [18]. We consider natural language to be a particularly useful modality for users to interact with robots, and have developed language-understanding methods for robot navigation using traditional techniques [19, 20, 21] and contemporary pre-trained models [22, 23]. Finally, robot tactile sensing and processing is still primitive compared to human touch; we have developed RL-based approaches that use tactile information to explore objects [24], recognize gestures [25], keep robots safe [26], and even to locomote on a snake robot [27, 28]. This work is supported by all the funding sources described above and a MathWorks Microgrant ($25K).

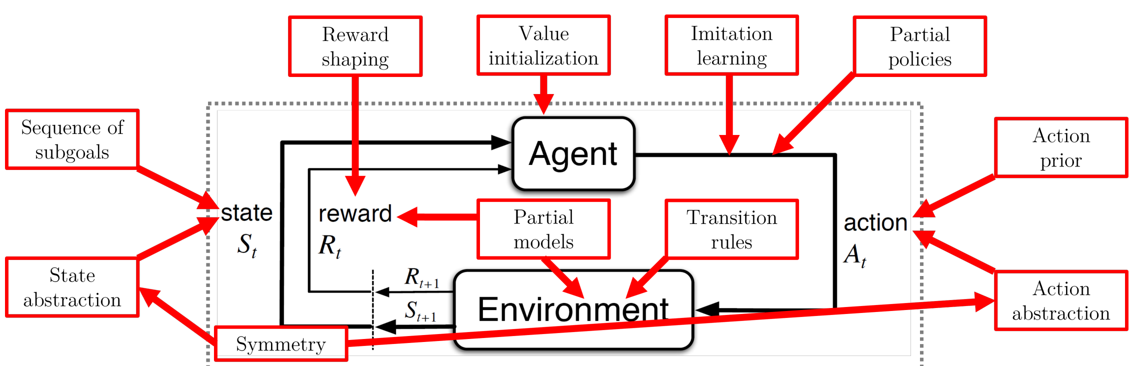
Future directions: Advice-guided reinforcement learning

The core theme that unites the research in my group is: using high-level information to improve the sample complexity and generalizability of RL and other decision-making approaches. So far, we have explored approaches to do this using state abstraction, symmetry, coarse maps, and demonstrations. We will continue pushing the boundary in these areas, including incorporating symmetry discovery in abstraction learning, handling more flexible coarse maps and diagrams that approach those humans typically use, and advancing high-level one-shot imitation learning. With the advent of large language models, we now also have a very rich, diverse, and occasionally incorrect source of advice for RL and robotics; we have performed preliminary work in this area [29, 30].

Beyond the types of high-level guidance described above, there are plenty of other MDP interfaces through which we might consider providing guidance to RL, as illustrated in the figure below. To lay the groundwork for future investigations, we need to answer several questions:

- How can RL algorithms use different forms of guidance, via exploration or directly?
- How effective is each form of guidance for speeding up learning and facilitating generalization?
- Are certain forms of guidance particularly prone to misspecification or incompleteness?
- What types and instances of guidance do users of all levels typically provide?

We hope to identify which types of advice are most important and relevant in various RL problems, which will guide our future exploration toward generalizable robotics and artificial intelligence.



Advice-guided RL: Interfaces for providing high-level guidance to Markov decision processes in RL.

References

[1] David Abel, Dilip Arumugam, Kavosh Asadi, Yuu Jinnai, Michael L. Littman, and Lawson L.S. Wong. State abstraction as compression in apprenticeship learning. In *AAAI Conference on Artificial Intelligence*, pages 3134–3142, 2019.

[2] Ondřej Bíža, Robert Platt, Jan-Willem van de Meent, and Lawson L.S. Wong. Learning discrete state abstractions with deep variational inference. In *Symposium on Advances in Approximate Bayesian Inference*, 2021.

[3] Linfeng Zhao, Xupeng Zhu, Lingzhi Kong, Robin Walters, and Lawson L.S. Wong. Integrating symmetry into differentiable planning with steerable convolutions. In *International Conference on Learning Representations*, 2023.

[4] Linfeng Zhao, Huazhe Xu, and Lawson L.S. Wong. Scaling up and stabilizing differentiable planning with implicit differentiation. In *International Conference on Learning Representations*, 2023.

[5] Linfeng Zhao, Hongyu Li, Taşkın Padır, Huaizu Jiang, and Lawson L.S. Wong. $E(2)$-equivariant graph planning for navigation. *IEEE Robotics and Automation Letters*, 9(4):3371–3378, 2024.

[6] Linfeng Zhao, Owen Howell, Xupeng Zhu, Jung Yeon Park, Zhewen Zhang, Robin Walters, and Lawson L.S. Wong. Equivariant action sampling for reinforcement learning and planning. In *Workshop on the Algorithmic Foundations of Robotics*, 2024.

[7] Ondřej Bíža, Thomas Kipf, David Klee, Robert Platt, Jan-Willem van de Meent, and Lawson L.S. Wong. Factored world models for zero-shot generalization in robotic manipulation. *arXiv preprint: 2202.05333*, 2022.

[8] Ondřej Bíža, Robert Platt, Jan-Willem van de Meent, Lawson L.S. Wong, and Thomas Kipf. Binding actions to objects in world models. In *International Conference on Learning Representations Workshop on the Elements of Reasoning: Objects, Structure, and Causality*, 2022.

[9] Linfeng Zhao, Lingzhi Kong, Robin Walters, and Lawson L.S. Wong. Toward compositional generalization in object-oriented world modeling. In *International Conference on Machine Learning*, pages 26841–26864, 2022.

[10] Dian Wang, Jung Yeon Park, Neel Sortur, Lawson L.S. Wong, Robin Walters, and Robert Platt. The surprising effectiveness of equivariant models in domains with latent symmetry. In *International Conference on Learning Representations*, 2023.

[11] Jung Yeon Park, Lawson L.S. Wong, and Robin Walters. Modeling dynamics over meshes with gauge equivariant nonlinear message passing. In *Neural Information Processing Systems*, pages 15277–15302, 2023.

[12] Chengguang Xu, Christopher Amato, and Lawson L.S. Wong. Robot navigation in unseen environments using coarse maps. In *IEEE International Conference on Robotics and Automation*, pages 2932–2938, 2024.

[13] Chengguang Xu, Christopher Amato, and Lawson L.S. Wong. Hierarchical robot navigation in novel environments using rough 2-D maps. In *Conference on Robot Learning*, pages 1971–1991, 2020.

[14] Linfeng Zhao and Lawson L.S. Wong. Learning to navigate in mazes with novel layouts using abstract top-down maps. In *Reinforcement Learning Conference*, volume 5, pages 2359–2372, 2024.

[15] Ondřej Bíža, Skye Thompson, Kishore Reddy Pagidi, Abhinav Kumar, Elise van der Pol, Robin Walters, Thomas Kipf, Jan willem van de Meent, Lawson L.S. Wong, and Robert Platt. One-shot imitation learning via interaction warping. In *Conference on Robot Learning*, pages 2519–2536, 2023.

[16] Jung Yeon Park and Lawson L.S. Wong. Robust imitation of a few demonstrations with a backwards model. In *Neural Information Processing Systems*, pages 19759–19772, 2022.

[17] Fan Xie, Alexander Chowdhury, M. Clara De Paolis Kaluza, Linfeng Zhao, Lawson L.S. Wong, and Rose Yu. Deep imitation learning for bimanual robotic manipulation. In *Neural Information Processing Systems*, pages 2327–2337, 2020.

[18] Ondřej Bíža, Dian Wang, Robert Platt, Jan-Willem van de Meent, and Lawson L.S. Wong. Action priors for large action spaces in robotics. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 205–213, 2021.

[19] Dilip Arumugam, Siddarth Karamcheti, Nakul Gopalan, Edward C. Williams, Mina Rhee, Lawson L.S. Wong, and Stefanie Tellex. Grounding natural language instructions to semantic goal representations for abstraction and generalization. *Autonomous Robots*, 43(2):449–468, 2019.

[20] Arthur Wandzel, Yoonseon Oh, Michael Fishman, Nishanth Kumar, Lawson L.S. Wong, and Stefanie Tellex. Multi-object search using object-oriented POMDPs. In *IEEE Int'l Conference on Robotics and Automation*, pages 7194–7200, 2019.

[21] Seth Pate, Wei Xu, Ziyi Yang, Maxwell Love, Siddarth Ganguri, and Lawson L.S. Wong. Natural language for human-robot collaboration: Problems beyond language grounding. In *AAAI Fall Symposium on Artificial Intelligence for Human-Robot Interaction*, 2021.

[22] Seth Pate and Lawson L.S. Wong. Indoor localization using vision and language. In *IEEE International Conference on Robot and Human Interactive Communication*, pages 1558–1564, 2023.

[23] Chengguang Xu, Christopher Amato, and Lawson L.S. Wong. Towards enhancing target detection for object goal navigation. In *IEEE International Conference on Robotics and Automation*, 2025. Currently under review.

[24] Shuo Jiang and Lawson L.S. Wong. Active tactile exploration using shape-dependent reinforcement learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 8995–9002, 2022.

[25] Shuo Jiang, Boce Hu, Linfeng Zhao, and Lawson L.S. Wong. Robot tactile gesture recognition based on full-body module e-skin. In *IEEE International Conference on Robotics and Automation*, 2025. Currently under review.

[26] Shuo Jiang, Adarsh Salagame, Alireza Ramezani, and Lawson L.S. Wong. Snake robot with tactile perception navigates on large-scale challenging terrain. In *IEEE International Conference on Robotics and Automation*, pages 5090–5096, 2024.

[27] Shuo Jiang and Lawson L.S. Wong. A hierarchical framework for robot safety using whole-body tactile sensors. In *IEEE International Conference on Robotics and Automation*, pages 8021–8028, 2024.

[28] Shuo Jiang, Adarsh Salagame, Alireza Ramezani, and Lawson L.S. Wong. Hierarchical RL-guided large-scale navigation of a snake robot. In *IEEE International Conference on Advanced Intelligent Mechatronics*, pages 1347–1352, 2024.

[29] Chengguang Xu, Hieu T. Nguyen, Christopher Amato, and Lawson L.S. Wong. Vision and language navigation in the real world via online visual language mapping. In *Conf. on Robot Learning Workshop on Language and Robot Learning*, 2023.

[30] Ondřej Bíža, Thomas Weng, Lingfeng Sun, Karl Schmeckpeper, Tarik Kelestemur, Yecheng Jason Ma, Robert Platt, Jan-Willem van de Meent, and Lawson L.S. Wong. On-robot reinforcement learning with goal-contrastive rewards. In *IEEE International Conference on Robotics and Automation*, 2025. Currently under review.