

---

# PDDLGYM: GYM ENVIRONMENTS FROM PDDL PROBLEMS

---

Tom Silver and Rohan Chitnis

MIT Computer Science and Artificial Intelligence Laboratory  
{tslvr, ronuchit}@mit.edu

February 26, 2020

## ABSTRACT

We present PDDL Gym, a framework that automatically constructs OpenAI Gym environments from PDDL domains and problems. Observations and actions in PDDL Gym are relational, making the framework particularly well-suited for research in relational reinforcement learning and relational sequential decision-making. PDDL Gym is also useful as a generic framework for rapidly building numerous, diverse benchmarks from a concise and familiar specification language. We discuss design decisions and implementation details, and also illustrate empirical variations between the 15 built-in environments in terms of planning and model-learning difficulty. We hope that PDDL Gym will facilitate bridge-building between the reinforcement learning community (from which Gym emerged) and the AI planning community (which produced PDDL). We look forward to gathering feedback from all those interested and expanding the set of available environments and features accordingly.

## 1 Introduction

The creation of benchmarks has often accelerated research progress in various subdomains of artificial intelligence [1, 2, 3]. In sequential decision-making tasks, tremendous progress has been catalyzed by benchmarks such as the environments in OpenAI Gym [4] and the planning tasks in the International Planning Competition (IPC) [5]. Gym defines a standardized way to interact with an environment, allowing easy comparison of various reinforcement learning algorithms. IPC provides a set of planning domains and problems written in Planning Domain Definition Language (PDDL) [6], allowing easy comparison of various symbolic planners.

In this work, we present PDDL Gym, an open-source framework<sup>1</sup> that combines elements of Gym and PDDL. **Concretely, PDDL Gym is a Python framework that automatically creates Gym environments from PDDL domain and problem files.** As with Gym, PDDL Gym allows for episodic, closed-loop interaction between the agent and the environment; the agent receives an observation from the environment and gives back an action, repeating this loop until the end of an episode. As in PDDL, PDDL Gym is fundamentally relational: observations are sets of ground relations over objects (e.g. `on(plate, table)`) and actions are templates ground with objects (e.g. `pick(plate)`). PDDL Gym is therefore particularly well-suited for relational learning and sequential decision-making research. See Figure 1 for renderings of some environments currently implemented in PDDL Gym, and Figure 2 for code examples.

The Gym API defines a hard boundary between the agent and the environment. In particular, the agent *only* interacts with the environment by taking actions and receiving observations. The environment implements a function `step` that advances the state given an action by the agent; `step` defines the transition model of the environment. Likewise, a PDDL domain encodes a transition model via its operators. However, in typical usage, PDDL is understood to exist entirely in the “mind” of the agent. A separate process is then responsible for transforming plans into executable actions that the agent can take in the world.

PDDL Gym defies this convention: in PDDL Gym, PDDL domains and problems lie firmly on the environment side of the agent-environment boundary. The environment uses the PDDL files to implement the `step` function that advances the state given an action. PDDL Gym is thus perhaps best understood as a *repurposing* of PDDL. Implementation-wise, this repurposing has subtle but important implications, discussed in (§2.2).

---

<sup>1</sup>Code available at <https://github.com/tomsilver/pddlgy>. Pull requests are welcome!

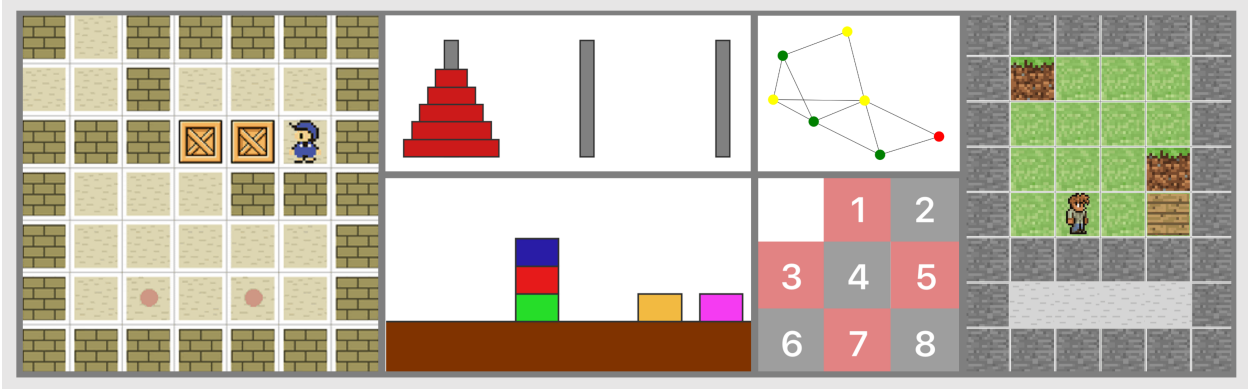


Figure 1: **Example environments implemented in PDDL Gym.** From top left: Sokoban, Hanoi, Blocks, Travelling Salesman, Slide Tile, and Crafting.

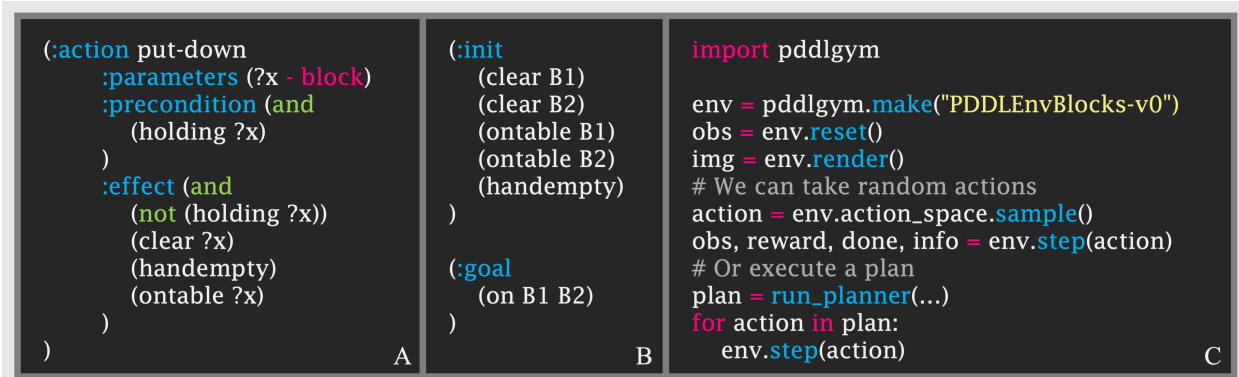


Figure 2: **PDDL Gym code examples.** A PDDL Gym environment is characterized by a PDDL domain file and a list of PDDL problem files. (A) One operator in the PDDL domain file for Blocks. (B) An excerpt of a simple PDDL problem file for Blocks. (C) After the PDDL domain and problem files have been used to register the environment with name “PDDLEnvBlocks-v0,” we can interact with this PDDL Gym environment in just a few lines of Python.

PDDL Gym serves three main purposes:

- (1) *Facilitate the creation of numerous, diverse benchmarks for sequential decision-making in relational domains.* PDDL Gym allows tasks to be defined in PDDL, automatically building a Gym environment from PDDL files. PDDL offers a compact symbolic language for describing domains, which might otherwise be cumbersome and repetitive to define directly via the Gym API.
- (2) *Bridge reinforcement learning and planning research.* PDDL Gym makes it easy for planning researchers and learning researchers to test their methods on the exact same benchmarks and develop techniques that draw on the strengths of both families of approaches. Furthermore, since PDDL Gym includes built-in domains and problems, it is straightforward to perform apples-to-apples comparisons without having to collect third-party code from disparate sources (see also [7]).
- (3) *Catalyze research on sequential decision-making in relational domains.* In our own research, we have found PDDL Gym to be very useful while studying exploration for lifted operator learning [8] and hierarchical goal-conditioned policy learning [9]. Other open research problems that may benefit from using PDDL Gym include relational reinforcement learning [10, 11, 12], learning symbolic descriptions of operators [10, 13, 14], discovering relational transition rules for efficient planning [15, 16], and learning lifted options [17, 18, 19, 20].

The rest of this paper is organized as follows. (§2) discusses the design decisions and implementation details underlying PDDL Gym. In (§3), we give an overview of the built-in PDDL Gym domains and provide basic empirical results to illustrate their diversity in terms of the difficulty of planning and learning. Finally, in (§4), we discuss avenues for extending and improving PDDL Gym.

## 2 Design and Implementation

The Gym API defines environments as Python classes with three essential methods: `__init__`, which initializes the environment; `reset`, which starts a new episode and returns an observation; and `step`, which takes an action from the agent, advances the current state, and returns an observation, reward, a Boolean indicating whether the episode is complete, and optional debugging information. The API also includes other minor methods, e.g., to handle rendering and random seeding. Finally, Gym environments are required to implement an `action_space`, which represents the space of possible actions, and an `observation_space`, which represents the space of possible observations. We next give a brief overview of PDDL files, and then we describe how action and observation spaces are defined in PDDL Gym. Subsequently, we move to a discussion of our implementation of the three essential methods.

### 2.1 Background: PDDL Domain and Problem Files

There are two types of PDDL files: domain files and problem files. A single *benchmark* is characterized by one domain file and multiple problem files.

A PDDL domain file includes *predicates* — named relations with placeholder variables such as (on ?x ?y) — and *operators*. An operator is composed of a name, a list of parameters, a first-order logic formula over the parameters describing the operator’s preconditions, and a first-order logic formula over the parameters describing the operator’s effects. The forms of the precondition and effect formulas are typically restricted depending on the version of PDDL. Early versions of PDDL only permit conjunctions of ground predicates [21]; later versions also allow disjunctions and quantifiers [22]. See Figure 2A for an example of a PDDL operator.

A PDDL problem file includes a set of *objects* (named entities), an *initial state*, and a *goal*. The initial state is a set of predicates ground with the objects. Any ground predicates not in the state are assumed to be false, following the closed-world assumption. The goal is a first-order logical formula over the objects (the form of the goal is limited by the PDDL version, like for operators’ preconditions and effects). Note that PDDL (and PDDL Gym) also allows objects and variables to be *typed*. See Figure 2B for a partial example of a PDDL problem file.

### 2.2 Observation and Action Spaces

In PDDL Gym, observations are sets of predicates grounded over the objects in the problem file. The observation space, then, represents the powerset of all possible ground predicates. This powerset is typically enormous; fortunately, it does not need to be explicitly computed. The `observation_space` can be viewed as a discrete space whose size is equal to the size of this powerset; since this space will be large, we expect that most algorithms for solving PDDL Gym tasks will not be sensitive to its size.

The action space for a PDDL Gym environment is one of the more subtle aspects of the overall framework. In classical AI planning, actions are typically equated with ground operators — operators whose parameters are bound to objects. However, in most PDDL domains, only some operator parameters are *free* (in terms of controlling the agent); the remaining parameters are included in the operator because they are part of the precondition/effect expressions, but can be derived from the current state or the choice of free parameters. PDDL typically makes no distinction between free and non-free parameters. For example, consider the operator for Sokoban shown in Figure 3A. This operator represents the rules for a player (?p) moving in some direction (?dir) from one cell (?from) to another cell (?to). In a real game of Sokoban, the only choice that an agent makes is what direction to move — only the ?dir parameter is free. ?from is defined by the agent’s location in the current state; ?to can be derived from ?from and the agent’s choice of ?dir.

To properly define the action space for a PDDL Gym environment, we must explicitly distinguish free parameters from non-free ones. One option is to require that operator parameters are all free. Non-free parameters could then be folded into the preconditions and effects using quantifiers [22]; see Figure 3B for an example. However, this is cumbersome and leads to clunky, deeply nested operators. Instead, we opt to introduce new predicates that represent operators, and whose variables are these operators’ free parameters. We then include these predicates in the preconditions of the respective operators; see Figure 3C for an example. Doing so requires only minimal changes to existing PDDL files and does not affect readability, but requires adding in domain knowledge about the agent-environment boundary. Note that this domain knowledge is equivalent to defining an action space, which is very commonly done in sequential decision-making and is not a strong assumption. The action space of a PDDL Gym environment is a discrete space over all possible groundings of the newly introduced predicates.

<pre> (action move :parameters (?p - thing ?from - loc              ?to - loc ?dir - dir) :precondition (and   (is-player ?p)   (at ?p ?from)   (clear ?to)   (move-dir ?from ?to ?dir) ) :effect (and   (not (at ?p ?from))   (not (clear ?to))   (at ?p ?to)   (clear ?from) ) ) </pre> <p style="text-align: center;">A</p>	<pre> (action move :parameters (?dir - dir) :precondition (and   (exists (?p - thing) (and     (is-player ?p)     (exists (?from - loc) (and       (at ?p ?from)       (exists (?to - loc) (and         (clear ?to)         (move-dir ?from ?to ?dir)       )     )   ) ) ) :effect (and   (forall (?p - thing) (     (when (forall (?from - loc) (       ...     )   ) ) ) ) ) </pre> <p style="text-align: center;">B</p>	<pre> (action move :parameters (?p - thing ?from - loc              ?to - loc ?dir - dir) :precondition (and   (move-action-selected ?dir)   (is-player ?p)   (at ?p ?from)   (clear ?to)   (move-dir ?from ?to ?dir) ) :effect (and   (not (at ?p ?from))   (not (clear ?to))   (at ?p ?to)   (clear ?from) ) ) </pre> <p style="text-align: center;">C</p>
--	---	--

Figure 3: **Explicating free parameters in PDDL operators.** PDDL operators traditionally conflate free and non-free parameters. For example, in a typical move operator for Sokoban (A), the free parameter `?dir` is included alongside non-free parameters. PDDL Gym must distinguish free parameters to properly define the action space. One option would be to require that all operator parameters are free, and introduce quantifiers in the operator body accordingly (B); however, this is cumbersome and leads to clunky operators, so we do *not* do this. Instead, we opt to introduce new predicates that represent operators, and whose variables are these operators’ free parameters (C). An example of such a new predicate is shown in yellow (`move-action-selected`).

## 2.3 Initializing and Resetting an Environment

A PDDL Gym environment is parameterized by a PDDL domain file and a list of PDDL problem files. For research convenience, each PDDL Gym environment is associated with a *test* version of that environment, where the domain file is identical but the problem files are different (for instance, they could encode more complicated planning tasks, to measure generalizability). During environment initialization, all of the PDDL files are parsed into Python objects; we use a custom PDDL parser for this purpose. When `reset` is called, a single problem instance is randomly selected.<sup>2</sup> The initial state of that problem instance is the state of the environment. For convenience, `reset` also returns (in the debugging information) paths to the PDDL domain and problem file of the current episode. This makes it easy for a user to run to a symbolic planner and execute resulting plans in the environment; see the README in the PDDL Gym Github repository for an example that uses FastForward [23].

## 2.4 Implementing step

The `step` method of a PDDL Gym environment takes in an action, updates the environment state, and returns an observation, reward, done Boolean, and debugging information. To determine the state update, PDDL Gym checks whether any PDDL operator’s preconditions are satisfied given the current state and action. Note that it is impossible to “accidentally match” to an undesired operator: each operator has a unique precondition as illustrated in Figure 3C, which is generated automatically based on the passed-in action. Since actions are distinct from operators (§2.2), this precondition satisfaction check is nontrivial; non-free parameters must be bound. To perform this check, we include in PDDL Gym a Python implementation of typed SLD resolution.<sup>3</sup> When no operator preconditions hold for a given action, the state remains unchanged by default. In some applications, it may be preferable to raise an error if no preconditions hold; the optional initialization parameter `raise_error_on_invalid_action` permits this behavior.

Rewards in PDDL Gym are sparse and binary. In particular, the reward is 1.0 when the problem goal is satisfied and 0.0 otherwise. Similarly, the done Boolean is True when the goal is reached and False otherwise. (In practice, a maximum episode length is often used.)

<sup>2</sup>Note that problem selection when resetting an episode is the only place where randomness is used in PDDL Gym.

<sup>3</sup>We also experimented with Prolog for inference, but found our own typed SLD resolution algorithm to be much faster, since Prolog is not natively typed.

Domain Name	Source	Rendering Included	Average FPS
<b>Baking</b>	Ours	No	2218
<b>Blocks</b>	[24]	Yes	3026
<b>Casino</b>	Ours	No	3781
<b>Crafting</b>	Ours	Yes	2195
<b>Depot</b>	[24]	No	315
<b>Doors</b>	[25]	Yes	699
<b>Elevator</b>	[24]	No	1026
<b>Ferry</b>	[26]	No	552
<b>Gripper</b>	[27]	Yes	2895
<b>Hanoi</b>	[27]	Yes	2409
<b>Meet-Pass</b>	[26]	No	4172
<b>Slide Tile</b>	[27]	Yes	1148
<b>Sokoban</b>	[24]	Yes	221
<b>TSP</b>	[27]	Yes	1196
<b>USA Travel</b>	Ours	No	1044

Table 1: **List of domains currently included in PDDL Gym, as of the date of preparation of this report (February 8, 2020).** For each environment, we report the original source of the PDDL files, whether or not we have implemented custom rendering, and the average frames per second (FPS).

## 2.5 Development Status

In terms of lines of code, the bulk of PDDL Gym is dedicated to PDDL file parsing and SLD resolution (used in `step`). We are continuing to develop both of these features so that a wider range of PDDL domains are supported. Aspects of PDDL 1.2 that are not yet supported by PDDL Gym include hierarchical typing, equality, quantifiers, and constant objects. Aspects of later PDDL versions, such as numerical fluents, are not supported. Our short-term objective is to provide full support for PDDL 1.2. We have found that a wide range of standard PDDL domains are already well-supported by PDDL Gym; see (§3) for an overview. We welcome requests for features and extensions, via either issues created on the Github page or email.

## 3 PDDL Gym by the Numbers

In this section, we start with an overview of the domains built into PDDL Gym, as of the date of preparation of this report (February 8, 2020). We then provide some experimental results that give insight into the variation between these domains, in terms of planning and model-learning difficulty. All experiments are performed on a single laptop with 32GB RAM and a 2.9GHz Intel Core i9 processor.

### 3.1 Overview of Environments

There are currently 15 domains built into PDDL Gym. Most of the domains are adapted from existing PDDL repositories; the remainder are ones we found to be useful benchmarks in our own research. We have implemented custom rendering for 8 of the domains (see Figure 1 for examples). Table 1 gives a list of all environments, their sources, and their average frames per second (FPS) calculated by executing a random policy for 1,000 episodes of horizon 10, with no rendering.

### 3.2 Variation in Environment Difficulty

We now provide some results illustrating the variation between the domains built into PDDL Gym. We examine two axes of variation: planning difficulty and difficulty of learning the transition model.

Figure 4 (left) illustrates the average time taken by FastForward [23] to find a plan in each of the environments, averaged across all problem instances. The results reveal a considerable range in planning time, with the most difficult domain (Depot, omitted from the plot for visual clarity) requiring two orders of magnitude more time than the simplest one (Baking). The results also indicate that the included domains are relatively “easy” from a modern planning perspective. However, even in these simple domains, there are many interesting challenges to be tackled, such as learning the true PDDL operators from interaction data, or defining good state abstractions amenable to learning.

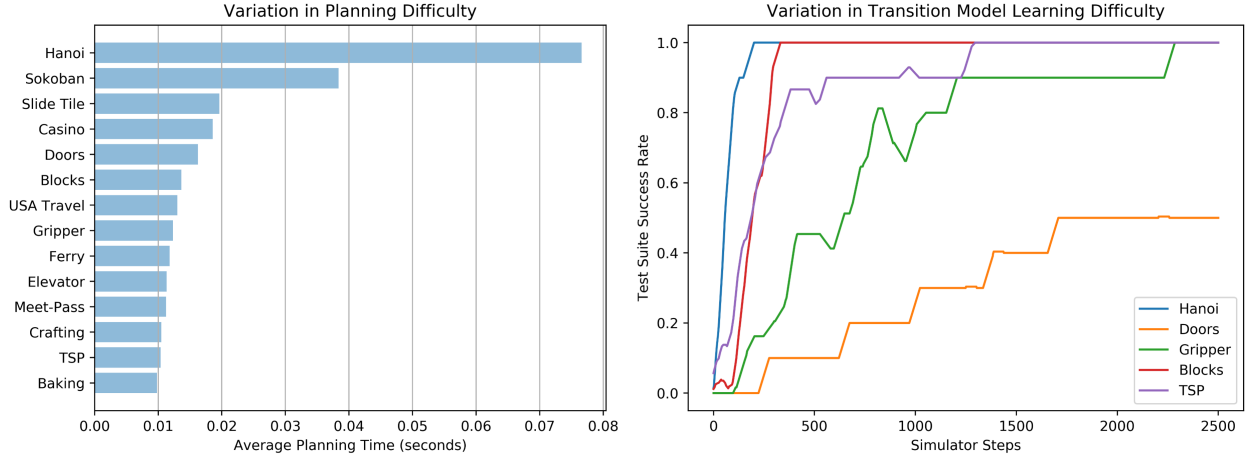


Figure 4: **Variation between PDDL Gym environments.** The PDDL domains and problems built into PDDL Gym vary considerably in terms of planning difficulty (left) and model learning difficulty (right). See text for details. Depot (1.16 seconds) is omitted from the left plot for clarity, but requires two orders of magnitude more time than Baking.

Figure 4 (right) provides insight into the difficulty of learning transition models in some of the environments. For each environment, an agent executes a random policy for episodes of horizon 25. The observed transitions are used to learn transition models, which are then used for planning on a suite of test problems. The fraction of test problems solved is reported as an indicator of the learned transition model. To learn the transition models, we use first-order logic decision tree (FOLDT) learning [28] (details will be described in forthcoming work). Five domains are visualized for clarity; among the remaining ones, several are comparable to the ones shown, but others, including Baking, Depot, and Sokoban, are difficult for our learning method: FOLDT learning is unable to find a model that fits the data in a reasonable amount of time. Of course, model-learning difficulty varies considerably with the learning method and the exploration strategy. We have implemented simple strategies here to show these results, but these avenues for future research are exactly the kind that we hope to enable with PDDL Gym.

## 4 Conclusion and Future Work

We have presented PDDL Gym, an open-source Python framework that automatically creates OpenAI Gym environments from PDDL domain and problem files. Our empirical results demonstrate considerable diversity among the built-in environments. We have already begun to use PDDL Gym in our own research into relational sequential decision-making and reinforcement learning. We also hope to interface PDDL Gym with other related open-source frameworks, particularly the PDDL collection and tools in `planning.domains` [7], so that a user can use PDDL Gym simply by specifying a URL pointing to a repository of PDDL files (along with some domain-specific information about free parameters). Another important avenue for future work is to incorporate stochasticity into the environments. Probabilistic PDDL [29] would be a natural starting point for such efforts.

We look forward to gathering feedback from the community and expanding the set of available environments and features accordingly.

## References

- [1] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [2] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv preprint arXiv:1804.07461*, 2018.
- [3] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.
- [4] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI gym. *arXiv preprint arXiv:1606.01540*, 2016.

- [5] Mauro Vallati, Lukas Chrpá, Marek Grześ, Thomas Leo McCluskey, Mark Roberts, Scott Sanner, et al. The 2014 international planning competition: Progress and trends. *Ai Magazine*, 36(3):90–98, 2015.
- [6] Drew McDermott, Malik Ghallab, Adele Howe, Craig Knoblock, Ashwin Ram, Manuela Veloso, Daniel Weld, and David Wilkins. Pddl-the planning domain definition language, 1998.
- [7] Christian Muise. Planning Domains. In *The 26th International Conference on Automated Planning and Scheduling - Demonstrations*, 2016.
- [8] Rohan Chitnis, Tom Silver, Joshua Tenenbaum, Leslie Pack Kaelbling, and Tomas Lozano-Perez. GLIB: Exploration via goal-literal babbling for lifted operator learning. *arXiv preprint arXiv:2001.08299*, 2020.
- [9] Tom Silver, Rohan Chitnis, Anurag Ajay, Josh Tenenbaum, and Leslie Pack Kaelbling. Learning skill hierarchies from predicate descriptions and self-supervision. In *AAAI GenPlan Workshop*, 2020.
- [10] Tobias Lang, Marc Toussaint, and Kristian Kersting. Exploration in relational domains for model-based reinforcement learning. *Journal of Machine Learning Research*, 13(Dec):3725–3768, 2012.
- [11] Sašo Džeroski, Luc De Raedt, and Kurt Driessens. Relational reinforcement learning. *Machine learning*, 43(1-2):7–52, 2001.
- [12] Prasad Tadepalli, Robert Givan, and Kurt Driessens. Relational reinforcement learning: An overview. In *Proceedings of the ICML-2004 workshop on relational reinforcement learning*, pages 1–9, 2004.
- [13] Eyal Amir and Allen Chang. Learning partially observable deterministic action models. *Journal of Artificial Intelligence Research*, 33:349–402, 2008.
- [14] Hanna M Pasula, Luke S Zettlemoyer, and Leslie Pack Kaelbling. Learning symbolic models of stochastic domains. *Journal of Artificial Intelligence Research*, 29:309–352, 2007.
- [15] Victoria Xia, Zi Wang, Kelsey Allen, Tom Silver, and Leslie Pack Kaelbling. Learning sparse relational transition models. *International Conference on Learning Representations*, 2019.
- [16] Tobias Lang and Marc Toussaint. Planning with noisy probabilistic relational rules. *Journal of Artificial Intelligence Research*, 39:1–49, 2010.
- [17] George Konidaris, Leslie Kaelbling, and Tomas Lozano-Perez. Constructing symbolic representations for high-level planning. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.
- [18] Martin Stolle and Doina Precup. Learning options in reinforcement learning. In *International Symposium on abstraction, reformulation, and approximation*, pages 212–223. Springer, 2002.
- [19] Doina Precup, Richard S Sutton, and Satinder Singh. Theoretical results on reinforcement learning with temporally abstract options. In *European conference on machine learning*, pages 382–393. Springer, 1998.
- [20] Nuttapon Chentanez, Andrew G Barto, and Satinder P Singh. Intrinsically motivated reinforcement learning. In *Advances in neural information processing systems*, pages 1281–1288, 2005.
- [21] Richard E Fikes and Nils J Nilsson. STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4):189–208, 1971.
- [22] Edwin PD Pednault. ADL: Exploring the middle ground between strips and the situation calculus. *Kr*, 89:324–332, 1989.
- [23] Jörg Hoffmann. FF: The fast-forward planning system. *AI magazine*, 22(3):57–57, 2001.
- [24] Malte Helmert. Pyperplan. <https://bitbucket.org/malte/pyperplan>, 2011.
- [25] George Konidaris and Andrew G Barto. Building portable options: Skill transfer in reinforcement learning. In *IJCAI*, volume 7, pages 895–900, 2007.
- [26] Colorado state university PDDL repository. <https://www.cs.colostate.edu/meps/repository.html>, 2002.
- [27] Soar group PDDL repository. <https://github.com/SoarGroup>, Accessed Jan 2020.
- [28] Hendrik Blockeel and Luc De Raedt. Top-down induction of first-order logical decision trees. *Artificial intelligence*, 101(1-2):285–297, 1998.
- [29] Håkan LS Younes and Michael L Littman. Ppddl1. 0: An extension to pddl for expressing planning domains with probabilistic effects. *Techn. Rep. CMU-CS-04-162*, 2:99, 2004.