From Data-based to Model-based AI: Representation Learning for Planning (RLEAP) *
(Short version)

Hector Geffner

29/8/2019

Abstract

Two of the main research threads in AI revolve around the development of **data-based learners** capable of inferring behavior and functions from experience and data, and **model-based solvers** capable of tackling well-defined but intractable models like SAT, classical planning, and Bayesian networks. Learners, and in particular deep learners, have achieved considerable success but result in black boxes that do not have the flexibility, transparency, and generality of their model-based counterparts. Solvers, on the the hand, require models which are hard to build by hand. RLEAP is aimed at achieving an integration of learners and solvers in the context of planning by addressing and solving **the problem of learning first-order planning representations from raw perceptions alone without using any prior symbolic knowledge.** The ability to construct first-order symbolic representations and using them for expressing, communicating, achieving, and recognizing goals is a main component of human intelligence and a fundamental, open research problem in AI. The success of RLEAP requires the development of radically new ideas and methods that will build on those of a number of related areas that include planning, learning, knowledge representation, combinatorial optimization and SAT. The approach to be pursued is based on a clear separation between learning the symbolic representations themselves, that is cast as a **combinatorial optimization problem**, and learning the interpretations of those representations, that is cast as a **supervised learning problem** from targets obtained from the first part. RLEAP will address both problems in the settings of planning and **generalized planning** where plans are general strategies. The project can make a significant difference in how **general, explainable, and trustworthy AI** can be understood and achieved.

1 Introduction

The project RLEAP aims to address and solve a fundamental research problem that is at the heart of the current split between **data-based learners** and **model-based reasoners (solvers)** in AI: the problem of **learning symbolic representations from raw perceptions**. The popularity of data-based learners over model-based solvers is that data is easily available but building models by hand is hard. Yet data-based learners lack the flexibility, transparency, and guarantees that are associated with model-based systems [LUTG17, Pea18, Dar18, Gef18]. By showing how to learn meaningful, symbolic models form raw perceptions alone, RLEAP aims to integrate the benefits of both. The context for learning representations is **planning** where representations play a key role in expressing, communicating, achieving, and recognizing goals [SA77, CL90, RG09, PC09, Gef13, SRBS16].

For an illustration of the **representation learning problem for planning** addressed by RLEAP, consider a range of 2D worlds where an agent must learn to achieve goals from scratch **from raw perceptions alone** (images) and **no prior symbolic knowledge.** The agent has to learn about the world in a flexible manner so that the knowledge gained for achieving goals in some worlds can

*Slightly edited version of Advanced ERC Project proposal, funded 1/10/2020–30/9/2025.

be reused to achieve related goals in other worlds. The goals are to be conveyed to the agent in a formal language whose grammar may be known to agent but whose symbols and meanings are not. The meanings of these symbols have to do with the objects and relations in the world which are not directly available to the agent who sees raw images only.

A version of the problem appears in the recent paper *BabyAI: A platform to study the sample efficiency of grounded language learning* by Yoshua Bengio and co-authors [CBBL+19]. In the paper, the authors describe a simulation platform for a class of 2D worlds featuring a number of objects and an agent that accepts goal instructions expressed in a context-free language. The agent learns to interpret and carry out the given goals from scratch by learning to maximize expected reward using **deep reinforcement learning** [MKS+15, SHM+16, SSS+17, SHS+18]. Among the conclusions, the authors state that "the methods scale and generalize poorly when it comes to learning tasks with a compositional structure." Hundreds of thousands of demonstrations are needed indeed to learn tasks which are trivial by human standards, and some simple tasks are not learned reliably at all.

The problem of representation learning for planning, while simple to describe and central to AI, is largely **unsolved**, and **current ideas and methods prove to be inadequate**. Indeed, two characteristics of deep reinforcement learning that have to do with its successes and its failures are its **ability** to deal with high dimensional perceptual spaces from scratch without prior knowledge, combined with its **inability** to use or produce such knowledge. Humans, on the other hand, excel at using prior knowledge when dealing with new tasks and at producing such knowledge when solving related tasks [LUTG17, Mar18a, Mar18b]. Certainly, the construction of reusable knowledge from experience (**transfer learning**) has been a central concern in reinforcement learning [TS11, Laz12] and in recent work in deep reinforcement learning [GDL+17, BBQ+19], but the semantic and conceptual gap between the **low level techniques** that are used, (neural network architectures and loss functions) and the **high-level representations** that are required (first-order representations involving objects and relations), remains just too large [AF18, TBF+18, FLBPP19, Asa19, GS19].

RLeap will develop the formulations and algorithms for showing how **first-order symbolic representations** involving objects and relations can be learned automatically from data **without using any prior symbolic knowledge**. Unlike current work in deep reinforcement learning, these representations will not be expected to emerge bottom-up from the learning process but will be forced top-down. We know indeed the structure of the first-order representations that are used in planning and the benefits that they have: they can be used to attain a variety of compound goals (**compositionality**), can be reused easily in a variety of problems (**transfer**), and can be queried at a high level of abstraction (**transparency**). There is thus no need to re-discover the structure of these representations nor to learn alternative ones that lack these properties. The challenge is to learn them from data.

2 Motivation

The current excitement about AI is the result of a number of breakthroughs in machine learning. [KSH12, GMH13, HZRS16, SHM+16, SSS+17]. Learners, like solvers, can be understood as programs that compute mappings from inputs x into outputs $f(x)$ by solving well-defined mathematical tasks [Gef14, Gef18]. In **deep learning** (DL) and **deep reinforcement learning** (DRL), training results in a function f that has a fixed structure given by a deep neural network [LBH15, GBC16] and a number of adjustable parameters. In DL, the input vector x may represent an image and the output $f(x)$, a classification label, while in DRL, the input x may represent the state of a game, and $f(x)$, the value of the state. For **solvers**, the input x represents a model instance, and the ouput $f(x)$, the solution to the instance x. Solvers have been developed for a variety of models that include constraint satisfaction problems (CSPs), SAT, answer set programs, Bayesian networks, classical planning, and various forms of probabilistic planning [Dec03, BHvM09, GKKS12, Pea88, GB13, Ber95].

The distinction between data-based learners and model-based solvers is reminiscent of the distinction between **Systems 1** and **2** in current psychologies theories of the human mind (Kahneman's Fast and Slow Thinking): the first referring to the **intuitive mind** that is fast, associative, unconscious, effortless, and parallel; the second to the **analytical mind** that is slow, deliberative, conscious, effortful, and serial [Kah11, ES13]. From this point of view, the learners deliver System 1 intelligence by producing fast black boxes that correspond to the learned functions f, while solvers exhibit System 2 intelligence by computing the outputs $f(x)$ for each given model instance x by reasoning.

A crucial difference between the human mind and current AI systems, however, is that **Systems 1 and 2 are tightly integrated** in the workings of the mind, while learners and solvers rarely talk to each other. This limitation explains why for example self-driving cars are unlikely to be deployed anytime soon: they are Systems 1 only, and as such, they cannot be trusted in open worlds where unexpected situations are bound to happen. By learning representations that enable reasoning on a case by case basis, RLeap will contribute to make AI systems from data that are able to integrate System 1 and System 2 intelligence.

3 Objectives (Summary)

The basic goal of the project is to address and solve the problem of **learning first-order symbolic representations from raw perceptions alone** in the context of planning where representations play a key role and their structure is known and provides the basic properties of **compositionality**, **reuse**, and **transparency**. This basic goal corresponds to Objective 1 below; Objectives 2 to 4 address closely related goals.

Objective 1: Learning representations for planning. A planning problem $P = \langle D, I\rangle$ combines a **first-order domain** D that contains action schemas, predicate symbols, and first-order atoms defining preconditions and effects, and **instance information** $I = \langle O, Init, Goal\rangle$ that encodes the relevant objects O and the sets of ground atoms that express the initial and goal conditions [McD00, HLMM19]. A planning problem P defines a directed graph $G(P)$ whose nodes n represent the states $s = s(n)$ over P, and whose edges (n, n') represent the state transitions. The states are represented by sets of ground atoms; namely, those that are true in the state.

For definining the basic representation learning problem, let an **image graph** G be a directed graph where each node n is associated with an observation that we take to be a **raw image** $v(n)$. The image graph can be obtained by sampling a large number of observed trajectories. We will assume that the graph is complete initially; namely, that all the possible observed trajectories are in the graph. The assumption is feasible if the problems used for training are small, i.e., if they involve hidden state spaces that are not too large, an idea that is compatible with *curriculum learning* [BLCW09].

The basic **representation learning problem** is to infer the **symbolic, first-order representation** of a set of planning instances $P_i = \langle D, I_i\rangle$, $i = 1, \ldots, n$ with a common domain D from **input data** given by a set of **image graphs** $G_1, \ldots, G_n$. As an example, consider 2D worlds that represent rectangular grids where an agent can move one unit at a time, collect keys one a time, and drop them. The observed trajectories can be generated with a simulator and a graphics engine. The learning problem is to infer the first-order representations $P_i = \langle D, I_i\rangle$ that account for the observations. This means **discovering** predicate symbols like loc^1, key^1, $adjacent^2$, $hold^1$, $handfree^0$, action schemas like $move^2$, $pickup^2$, $drop^2$, and so on, or equivalent representations, **from raw perceptions alone.**

A key assumption is that different states give rise to different images. Partial observability and non-deterministic actions will also be considered on top of this basic setting, taking advantage that the languages used for planning in the richer settings are built on top of this basic language [YLWA05, CCO+12]. Noisy observations will be addressed too. The basic problem, however, is central and challenging enough, and far from being solved.

Objective 2: Learning representations for generalized planning. Generalized planning studies the methods for expressing and obtaining plans that will solve not just one planning instance but multiple instances. For example, a **general plan** for solving **any instance** of the domain where a key is to be picked up and delivered to a target location is simple: the agent has to go to the key, pick it up, go to the target location, and drop it. The challenge is to obtain such general plans automatically [SIZ08, BPG09, SIZ11, HD11, BL16, SAJJ16, IM19]. In recent years, it has been shown by the PI and others that such general plans can be derived through reductions and off-the-shelf planners from a **generalized model** that provides a common abstraction of the instances to be solved [BG18b, BFG19]. Objective 2 is **learning such generalized models directed from perceptual data.** The objective ties closely with knowledge representation on the one hand, and learning on the other. Indeed, learning approaches are not aimed at finding plans for single instances but general plans [CBBL$^+$19, FLBPP19, GS19].

Objective 3: Learning hierarchical representations. In order to compute plans it is often necessary or convenient to plan at different levels of abstractions, with the constraints among the different levels playing a crucial role. Indeed, a high-level plan is useless if it can't be brought down to the lowest level for execution [BY91]. In planning and reinforcement learning, hierarchies and high level actions or options have been defined mostly **by hand** [HTD90, NAI$^+$03, GA15, BAH19, SPS99]. In spite of many efforts [KB07, MRW07, BHP17, MBB17], **the key problem** that remains open and will be addressed in RLeap is how to discover crisp, symbolic hierarchical representations automatically, constructing successive layers of abstractions from the bottom up. The new elements that will be exploited are the intimate connection between hierarchical planning and generalized planning, as **hierarchical plans are hierarchical general plans**, and recent work by the PI and team that shows how abstractions for generalized planning can be obtained automatically from a first-order symbolic representation of the domain [BG18b, BFG19].

Objective 4: Theory of representations for planning and learning. There are two implicit assumptions in the project: one that the target representations to be learned are simple (have a low dimensionality), the other, that planning with these representations is simple as well (low polynomial time). These assumptions hold well in planning where 1) domains involve a bounded number of action schemas and predicate symbols, 2) planners scale up well [BG01, HN01, RW10, LG17a], in spite of the worst-case complexity results [Byl94]. Some of the new algorithms are indeed exponential in a **width** parameter that is small and bounded for most planning domains when goals are single atoms [LG12, LG17b, LG17a, FRLG17]. The formal proofs that establish that a domain has bounded width actually uncovers domain **features** (numerical state functions) that may shed light on two apparently unrelated open problems in **generalized planning** and **reinforcement learning**: what are the features required for producing general plans in a given domain, and what are the features that make linear function approximations work in a given domain [SPLC16, BDD$^+$19, BGP19]. Objective 4 is about developing a theory that formally relates the **features** that appear in three contexts, which may have much in common: **width analysis**, **linear function approximation in RL**, and **generalized planning**.

4 Feasibility and Novelty (Summary)

The **feasibility** and **novelty** of RLeap rest on two main premises and the way in which we will formulate them mathematically and computationally. First, that the language for extracting, using, reusing, and composing knowledge is the language of first-order symbolic representations, and that it is not necessary nor convenient to learn the structure of such languages from scratch. Second, that it is precisely the structure of such languages that provides the **strong structural priors** that make the learning of crisp representations feasible and data efficient.

RLeap will not follow deep learning approaches in assuming that the target representations emerge from the learning process through the use of suitable neural architectures (e.g., attention mechanisms) and loss functions (e.g., that penalize entanglement). Instead, RLeap will make first-order representations the explicit target of the learning process, and by doing so, it will decompose the representation learning problem in two: a **representation discovery problem**, that is a purely combinatorial problem, and a **semantic interpretation problem**, that is a supervised learning problem with targets obtained from the first part. Going back to the example above: discovering the action schemas with the predicate symbols loc^1, key^1, $adjacent^2$, $hold^1$, $handfree^0$ or equivalent ones from the input graphs is the representation discovery problem. Learning the functions that provide the denotation of such logical symbols in the images is the semantic interpretation problem.

The **representation discovery problem** is **combinatorial** because the number of possible domains given a bound on the number of action schemas, predicate symbols, and their arities, is bounded. The values of these parameters are bounded and small, and do not grow with the size of the instances. The problem of learning the simplest planning instances $P_i = \langle D, I_i \rangle$ that account for a number of input image graphs G_i, $i = 1, \ldots, m$ can then be cast and solved as a **combinatorial optimization** problem. An instance P_i accounts for the image graph G_i if the graph $G(P_i)$ associated with P_i is structurally equivalent (isomorphic) to the plain graph G_i; i.e., if the graph G_i, leaving the images aside, is **generated** by the planning instance P_i.

A **key property** of this view is that the first-order symbolic representations that are learned from the input graphs $G_1, \ldots, G_m$, i.e., the action schemas and the predicate symbols, do not depend on the **raw images** $v(n)$ associated with the nodes n but on the structure of the graphs. This means that the way in which objects are displayed on the images may change but the resulting representation will not. This is very different from works where the representations are obtained from auto-encoders and hence are low dimensional representations of the images [KW14, AF18, Asa19]. In the proposed formulation, **the symbolic representations do not provide a compact encoding of the images but of the structure of the state space.**

The images play a key role in the **semantic interpretation problem** that must yield the **functions** that provide the denotation of the learned predicate symbols in the images; namely, the tuples of objects that satisfy the predicate in the image. The problem is similar to the **semantic scene interpretation** problem [NM08, KZG+17, DSG17] where a raw image must be mapped into a logical formula that describes the contents of the image, in our case, the set (conjunction) of ground atoms $p(c_1, \ldots, c_k)$ that are true in the image for each of the learned predicate symbols p. In our setting, however, the semantic interpretation problem does **not** require labeled examples (**supervision**) as the solution of the representation discovery problem matches the nodes n in the graphs G_i with states $s(n)$ in the graphs $G(P_i)$, so that every image $v(n)$ is associated with a set of atoms $s(n)$.

The semantic interpretation problem is not trivial but it is much simpler than the full representation learning problem itself, and can be addressed through a principled combination of **fast convolutional object detectors** [RDGF16, RF17], **relation classifiers** [ZK15, SRB+17, SYZ+18] and **combinatorial optimizers** [BJM+18, FM06, ABL13, MML14], that yield the most likely scene interpretation in terms of the learned predicate symbols, using extra logical constraints and invariants that can be obtained from the learned planning instances P_i [Hel09, Rin17].

These ideas pertain mainly to Objectives 1 and 2. Objective 3 is about constructing new symbolic representations from given (learned) symbolic representations, and so is the problem of aligning these symbolic representations with given goal instructions. Objective 4 is about elaborating a theory of tractable representations for plannning and learning. The work in these two objectives will benefit from the expertise of PI and his team that introduced the notion of sound abstractions for generalized planning [BDGGR17, BG18b, BFG19], and the notion of width [LG12, LG17a, LG17b, FRLG17].

RLeap aims at methods that are domain-independent and that can be evaluated empirically in the way that planners and SAT solvers are tested in competitions. The methods to be developed are to be applicable to existing simulation platforms like ALE, for the Atari games [BNVB13], and GVG-AI, for general video-games [PLST+15], but for getting there, we will move one step at a time. We are not just interested in performance and coverage, but also in understanding. There are indeed model-based RL approaches for some of these domains based on the prediction of screens and rewards [OGL+15, KBM+19], but none that builds meaningful first-order symbolic representations from the screen that are used to plan.

5 State of the art. Related Research

AI systems that are general, explainable, and trustworthy require System 1 and System 2 intelligences tightly integrated, yet practically no current system achieves an integration of this sort. The **AlphaZero** program [SSS+17, SHS+18], that achieved suprahuman performance in Go and Chess, integrates a deep reinforcement learner [MKS+15] and a Monte Carlo Tree Search (MCTS) planner [KS06, CWH+08], but the learning is about the level of play, not about the model that is fixed and known in advance. More general AI systems will have to **learn models from data**, and moreover, the learned models must be structured in terms of objects and relations to facilitate reuse, transparency, and compositionality. RLeap is aimed at this challenge in the context of planning where **first-order representations** are known to provide these benefits [McD00, HLMM19]. These representations for planning, however, are written by hand. The main challenge addressed by RLeap is to **learn them from raw perceptions alone.**

Model learning is the goal of **model-based reinforcement learning** where the model parameters of a Markov Decision Processes (state transition probabilities and rewards) are learned by active exploration [SB98, BT03]. In the standard setting, the states are given and assumed to be fully observable but what is learned in one model does not transfer easily to other models. In particular, no high-level symbolic representations are learned. An exception is the work on **object-oriented MDPs** that starts and refines a **first-order symbolic representation** made up of objects, types, and relations [DCL08, HMT15]. Similar work has been done in classical planning [YWJ07, AJOR19]. In the two cases, however, the first-order symbolic models are obtained using predicate symbols with a known meaning that are given. **Inductive logic programming** approaches have this form as well obtaining symbolic representations for new concepts in the form of logic programs from given symbolic predicates and a number of positive and negative samples [MDR94, DRK08]. Variations of these ideas have been used in **hybrid learning schemes** that integrate symbolic and deep learning [MDK+18, KP18, SG16, XZF+18, GGL+19, EG18] and in **relational reinforcement learning** for obtaining general policies [Kha99, MG04, FYG04, DDRD01, KODR04]. A **general policy** is a policy that can solve problems that involve different sets of objects, configurations, and state spaces. More recently, **deep learning** methods have been used to obtain such policies, but once again, starting with the first-order (PDDL) symbolic representations of the domains [TTTX18, BdBMS19, IFT18, BG+18a]. A model-based approach for obtaining general policies from the same representations is developed in [BG18b, BFG19]. A related formulation finds abstract symbolic representations from low level symbolic representations and a given set of high-level options [KKLP18].

Deep reinforcement learning (DRL) methods have emerged as the main approach capable of generating general policies over high-dimensional perceptual spaces **without using any prior symbolic knowledge** [MKS+15, SHM+16, SSS+17]. Yet by not using or constructing first-order symbolic representations, DRL methods have not managed to obtain the benefits of transparency, reuse, and compositionality [CBBL+19, Mar18a, LB17]. Recent work in deep symbolic relational reinforcement learning [GS19] attempts to account for objects and relations through the use of attention mechanisms and suitable loss functions, but the **semantic and conceptual gap** between these **low level techniques** and the **high-level representations** that are required remains just too large. Something similar occurs with work aimed at learning low dimensional representations that disentangle the factors of variations in the data [TBF+18, FLBPP19]. The first-order representation used in planning have indeed a low dimensionality given by a small number of action schemas, predicate symbols, and atoms, but it is not clear how such representations can emerge bottom up from current architectures.

A recent deep learning approach infers compact representations for planning from raw images alone using a class of **variational autoencoders** where the representations provide a low dimensional encoding of the images [AF18]. Follow up work has extended this approach for producing first-order planning representations as well [Asa19]. It is far from clear, however, that the high level representations that are required are compact encodings of images. The standard PDDL planning files do not actually codify images but some of the structural relations that appear in the images. Moreover, it may be difficult to get crisp, compact symbolic representations when representations are forced to encode the images themselves.

RLeap departs from existing approaches to representation learning by assuming that the target representations encode the **structure of the state space**, not the way in which the states are visualized. In other words, if the images used to display the states are changed, while keeping the condition that different hidden states are displayed differently, the perceived state space and the first-order representations that are obtained from it, will not change. What will change in that case is the interpretation of the symbols learned in the images. We draw for this on the distinction between **symbols** and their **denotation** in first-order logic [Men09, HR04]. A first-order semantic interpretation provides a denotation to every (non-logical symbol) so that every (closed) first-order term and formula can be evaluated. In particular, terms t denote objects from the interpretation domain, and predicate symbols p denote boolean functions. The denotation of an atom like $p(t_1, \ldots, t_k)$ is *true* if the denotation of p maps the tuple of objects denoted by the terms t_1 to t_k into *true*. The *extension* of the predicate symbol p in the interpretation is defined as the set of object tuples that are mapped to *true*, and it is an alternative representation of the denotation function of p.

6 Conclusion

RLeap is aimed at a concrete scientific problem and some of its ramifications: learning structural symbolic representations for planning from images without using prior symbolic knowledge. The problem is central for developing flexible, transparent, and trusworthy AI systems, but it is largely unsolved, with current ideas and methods proving to be inadequate. The challenge is big but we bring to the project a number of concrete, promising ideas and formulations, many of them introduced by the PI and his team. The potential gains are important, affecting the way AI systems are built, used, verified, and queried, while providing an integration of data-based learners and model-based solvers with their System 1 (reactive) and System 2 (deliberative) capabilities.

References

[ABL13] Carlos Ansótegui, Maria Luisa Bonet, and Jordi Levy. Sat-based maxsat algorithms. *Artificial Intelligence*, 196:77–105, 2013.

[AF18] Masataro Asai and Alex Fukunaga. Classical planning in deep latent space: Bridging the subsymbolic-symbolic boundary. In *AAAI*, 2018.

[AJOR19] Diego Aineto, Sergio Jiménez, Eva Onaindia, and Miquel Ramírez. Model recognition as planning. In *Proc. ICAPS*, pages 13–21, 2019.

[Asa19] Masataro Asai. Unsupervised grounding of plannable first-order logic representation from images. In *Proc. ICAPS*, 2019.

[BAH19] Pascal Bercher, Ron Alford, and Daniel Höller. A survey on hierarchical planning–one abstract idea, many concrete realizations. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019). IJCAI*, 2019.

[BBQ+19] André Barreto, Diana Borsa, John Quan, Tom Schaul, David Silver, Matteo Hessel, Daniel Mankowitz, Augustin Žídek, and Remi Munos. Transfer in deep reinforcement learning using successor features and generalised policy improvement. *arXiv preprint arXiv:1901.10964*, 2019.

[BdBMS19] Thiago P Bueno, Leliane N de Barros, Denis D Mauá, and Scott Sanner. Deep reactive policies for planning in stochastic nonlinear domains. In *AAAI*, volume 33, pages 7530–7537, 2019.

[BDD$^+$19] Marc G Bellemare, Will Dabney, Robert Dadashi, Adrien Ali Taiga, Pablo Samuel Castro, Nicolas Le Roux, Dale Schuurmans, Tor Lattimore, and Clare Lyle. A geometric perspective on optimal representations for reinforcement learning. *arXiv preprint arXiv:1901.11530*, 2019.

[BDGGR17] Blai Bonet, Giuseppe De Giacomo, Hector Geffner, and Sasha Rubin. Generalized planning: Non-deterministic abstractions and trajectory constraints. In *Proc. IJCAI*, 2017.

[Ber95] Dimitri Bertsekas. *Dynamic Programming and Optimal Control, Vols 1 and 2.* Athena Scientific, 1995.

[BFG19] Blai Bonet, Guillem Francès, and Hector Geffner. Learning features and abstract actions for computing generalized plans. In *Proc. AAAI*, 2019.

[BG01] Blai Bonet and Hector Geffner. Planning as heuristic search. *Artificial Intelligence*, 129(1–2):5–33, 2001.

[BG$^+$18a] Aniket Nick Bajpai, Sankalp Garg, et al. Transfer of deep reactive policies for mdp planning. In *Advances in Neural Information Processing Systems*, pages 10965–10975, 2018.

[BG18b] Blai Bonet and Hector Geffner. Features, projections, and representation change for generalized planning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 4667–4673. AAAI Press, 2018.

[BGP19] Bahram Behzadian, Soheil Gharatappeh, and Marek Petrik. Fast feature selection for linear value function approximation. In *ICAPS*, 2019.

[BHP17] Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In *AAAI*, 2017.

[BHvM09] Armin Biere, Marijn Heule, and Hans van Maaren. *Handbook of satisfiability.* IOS press, 2009.

[BJM$^+$18] Fahiem Bacchus, Matti Juhani Järvisalo, Ruben Martins, et al. Maxsat evaluation 2018. 2018.

[BL16] Vaishak Belle and Hector J. Levesque. Foundations for generalized planning in unbounded stochastic domains. In *KR*, pages 380–389, 2016.

[BLCW09] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proc. ICML*, pages 41–48, 2009.

[BNVB13] Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, 47:253–279, 2013.

[BPG09] Bonet Bonet, Hector Palacios, and Hector Geffner. Automatic derivation of memoryless policies and finite-state controllers using classical planners. In *Proc. ICAPS-09*, pages 34–41, 2009.

[BT03] Ronen I. Brafman and Moshe Tennenholtz. R-max-a general polynomial time algorithm for near-optimal reinforcement learning. *The Journal of Machine Learning Research*, 3:213–231, 2003.

[BY91] Fahiem Bacchus and Qiang Yang. The downward refinement property. In *IJCAI*, pages 286–293, 1991.

[Byl94] Thomas Bylander. The computational complexity of STRIPS planning. *Artificial Intelligence*, 69:165–204, 1994.

[CBBL+19] Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, and Yoshua Bengio. Babyai: A platform to study the sample efficiency of grounded language learning. In *ICLR*, 2019.

[CCO+12] Amanda Coles, Andrew Coles, Angel García Olaya, Sergio Jiménez, Carlos Linares López, Scott Sanner, and Sungwook Yoon. A survey of the seventh international planning competition. *AI Magazine*, 33(1):83–88, 2012.

[CL90] Philip R Cohen and Hector J Levesque. Intention is choice with commitment. *Artificial intelligence*, 42(2-3):213–261, 1990.

[CWH+08] Guillaume M JB Chaslot, Mark HM Winands, H JAAP VAN DEN HERIK, Jos WHM Uiterwijk, and Bruno Bouzy. Progressive strategies for monte-carlo tree search. *New Mathematics and Natural Computation*, 4(03):343–357, 2008.

[Dar18] Adnan Darwiche. Human-level intelligence or animal-like abilities? *Communications of the ACM*, 61(10):56–67, 2018.

[DCL08] Carlos Diuk, Andre Cohen, and Michael L Littman. An object-oriented representation for efficient reinforcement learning. In *Proceedings of the 25th international conference on Machine learning*, pages 240–247, 2008.

[DDRD01] Sašo Džeroski, Luc De Raedt, and Kurt Driessens. Relational reinforcement learning. *Machine learning*, 43(1-2):7–52, 2001.

[Dec03] Rina Dechter. *Constraint Processing*. Morgan Kaufmann, 2003.

[DRK08] Luc De Raedt and Kristian Kersting. Probabilistic inductive logic programming. In *Probabilistic Inductive Logic Programming*, pages 1–27. Springer, 2008.

[DSG17] Ivan Donadello, Luciano Serafini, and Artur D'Avila Garcez. Logic tensor networks for semantic image interpretation. In *Proc. IJCAI*, pages 1596–1602, 2017.

[EG18] Richard Evans and Edward Grefenstette. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018.

[ES13] Jonathan Evans and Keith Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3), 2013.

[FLBPP19] Vincent François-Lavet, Yoshua Bengio, Doina Precup, and Joelle Pineau. Combined reinforcement learning via abstract representations. In *Proc. AAAI*, volume 33, pages 3582–3589, 2019.

[FM06] Zhaohui Fu and Sharad Malik. On solving the partial max-sat problem. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 252–265. Springer, 2006.

[FRLG17] Guillem Francès, Miquel Ramírez, Nir Lipovetzky, and Hector Geffner. Purely declarative action representations are overrated: Classical planning with simulators. In *Proc. IJCAI*, 2017.

[FYG04] Alan Fern, SungWook Yoon, and Robert Givan. Approximate policy iteration with a policy language bias. In *Advances in neural information processing systems*, pages 847–854, 2004.

[GA15] Ilche Georgievski and Marco Aiello. Htn planning: Overview, comparison, and beyond. *Artificial Intelligence*, 222:124–156, 2015.

[GB13] Hector Geffner and Blai Bonet. *A concise introduction to models and methods for automated planning*. Morgan & Claypool, 2013.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[GDL+17] Abhishek Gupta, Coline Devin, YuXuan Liu, Pieter Abbeel, and Sergey Levine. Learning invariant feature spaces to transfer skills with reinforcement learning. 2017.

[Gef13] Hector Geffner. Computational models of planning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(4):341–356, 2013.

[Gef14] Hector Geffner. Artificial intelligence: From programs to solvers. *AI Communications*, 2014.

[Gef18] Hector Geffner. Model-free, model-based, and general intelligence. In *IJCAI*, 2018.

[GGL+19] Artur d'Avila Garcez, Marco Gori, Luis C Lamb, Luciano Serafini, Michael Spranger, and Son N Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv preprint arXiv:1905.06088*, 2019.

[GKKS12] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. *Answer set solving in practice*. Morgan & Claypool, 2012.

[GMH13] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.

[GS19] Marta Garnelo and Murray Shanahan. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences*, 29:17–23, 2019.

[HD11] Yuxiao Hu and Giuseppe De Giacomo. Generalized planning: Synthesizing plans that work for multiple environments. In *IJCAI*, pages 918–923, 2011.

[Hel09] Malte Helmert. Concise finite-domain representations for pddl planning tasks. *Artificial Intelligence*, 173(5-6):503–535, 2009.

[HLMM19] Patrik Haslum, Nir Lipovetzky, Daniele Magazzeni, and Christian Muise. *An Introduction to the Planning Domain Definition Language*, volume 13. Morgan & Claypool, 2019.

[HMT15] David Ellis Hershkowitz, James MacGlashan, and Stefanie Tellex. Learning propositional functions for planning and reinforcement learning. In *2015 AAAI Fall Symposium Series*, 2015.

[HN01] Joerg Hoffmann and Bernhard Nebel. The FF planning system: Fast plan generation through heuristic search. *Journal of Artificial Intelligence Research*, 14:253–302, 2001.

[HR04] Michael Huth and Mark Ryan. *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press, 2004.

[HTD90] James A Hendler, Austin Tate, and Mark Drummond. AI planning: Systems and techniques. *AI magazine*, 11(2):61–61, 1990.

[HZRS16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016.

[IFT18] Murugeswari Issakkimuthu, Alan Fern, and Prasad Tadepalli. Training deep reactive policies for probabilistic planning problems. In *ICAPS*, 2018.

[IM19] León Illanes and Sheila A McIlraith. Generalized planning via abstraction: arbitrary numbers of objects. In *Proc. AAAI*, 2019.

[Kah11] Daniel Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, 2011.

[KB07] George Konidaris and Andrew G Barto. Building portable options: Skill transfer in reinforcement learning. In *IJCAI*, volume 7, pages 895–900, 2007.

[KBM+19] Lukasz Kaiser, Mohammad Babaeizadeh, Piotr Milos, Blazej Osinski, Roy H Campbell, Konrad Czechowski, Dumitru Erhan, Chelsea Finn, Piotr Kozakowski, Sergey Levine, et al. Model-based reinforcement learning for atari. *arXiv preprint arXiv:1903.00374*, 2019.

[Kha99] Roni Khardon. Learning action strategies for planning domains. *Artificial Intelligence*, 113(1-2):125–148, 1999.

[KKLP18] George Konidaris, Leslie Pack Kaelbling, and Tomas Lozano-Perez. From skills to symbols: Learning symbolic representations for abstract high-level planning. *Journal of Artificial Intelligence Research*, 61:215–289, 2018.

[KODR04] Kristian Kersting, Martijn Van Otterlo, and Luc De Raedt. Bellman goes relational. In *Proceedings of the twenty-first international conference on Machine learning*, page 59. ACM, 2004.

[KP18] Seyed Mehran Kazemi and David Poole. Relnn: A deep neural model for relational learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[KS06] L. Kocsis and C. Szepesvári. Bandit based Monte-Carlo planning. In *Proc. ECML-2006*, pages 282–293, 2006.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012.

[KW14] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *ICLR*, 2014.

[KZG+17] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1):32–73, 2017.

[Laz12] Alessandro Lazaric. Transfer in reinforcement learning: a framework and a survey. In *Reinforcement Learning*, pages 143–173. Springer, 2012.

[LB17] Brenden M Lake and Marco Baroni. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. *arXiv preprint arXiv:1711.00350*, 2017.

[LBH15] Yann LeCun, Yoshua Bengio, and Goeffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[LG12] Nir Lipovetzky and Hector Geffner. Width and serialization of classical planning problems. In *Proc. ECAI*, 2012.

[LG17a] Nir Lipovetzky and Hector Geffner. Best-first width search: Exploration and exploitation in classical planning. In *Proc. AAAI*, 2017.

[LG17b] Nir Lipovetzky and Hector Geffner. A polynomial planning algorithm that beats lama and ff. *Proc. ICAPS*, 2017.

[LUTG17] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum, and Samuel J Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40, 2017.

[Mar18a] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[Mar18b] Gary Marcus. Innateness, AlphaZero, and artificial intelligence. *arXiv preprint arXiv:1801.05667*, 2018.

[MBB17] Marios C Machado, Marc G Bellemare, and Michael Bowling. A laplacian framework for option discovery in reinforcement learning. In *ICML*, pages 2295–2304, 2017.

[McD00] Drew McDermott. The 1998 AI Planning Systems Competition. *Artificial Intelligence Magazine*, 21(2):35–56, 2000.

[MDK+18] Robin Manhaeve, Sebastijan Dumancic, Angelika Kimmig, Thomas Demeester, and Luc De Raedt. Deepproblog: Neural probabilistic logic programming. In *Advances in Neural Information Processing Systems*, pages 3749–3759, 2018.

[MDR94] Stephen Muggleton and Luc De Raedt. Inductive logic programming: Theory and methods. *The Journal of Logic Programming*, 19:629–679, 1994.

[Men09] Elliott Mendelson. *Introduction to mathematical logic*. Chapman and Hall/CRC, 2009.

[MG04] Mario Martín and Hector Geffner. Learning generalized policies from planning examples using concept languages. *Applied Intelligence*, 20(1):9–19, 2004.

[MKS+15] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[MML14] Ruben Martins, Vasco Manquinho, and Inês Lynce. Open-WBO: A modular MaxSAT solver. In *Proc. SAT*, pages 438–445, 2014.

[MRW07] Bhaskara Marthi, Stuart J Russell, and Jason Andrew Wolfe. Angelic semantics for high-level actions. In *ICAPS*, pages 232–239, 2007.

[NAI+03] Dana S Nau, Tsz-Chiu Au, Okhtay Ilghami, Ugur Kuter, J William Murdock, Dan Wu, and Fusun Yaman. Shop2: An htn planning system. *Journal of artificial intelligence research*, 20:379–404, 2003.

[NM08] Bernd Neumann and Ralf Möller. On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82–101, 2008.

[OGL+15] Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard L Lewis, and Satinder Singh. Action-conditional video prediction using deep networks in atari games. In *Advances in neural information processing systems*, pages 2863–2871, 2015.

[PC09] Giovanni Pezzulo and Cristiano Castelfranchi. Intentional action: from anticipation to goal-directed behavior. *Psychological Research*, 2009.

[Pea88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, 1988.

[Pea18] Judea Pearl. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*, 2018.

[PLST+15] Diego Perez-Liebana, Spyridon Samothrakis, Julian Togelius, Tom Schaul, Simon M Lucas, Adrien Couëtoux, Jerry Lee, Chong-U Lim, and Tommy Thompson. The 2014 general video game playing competition. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(3):229–243, 2015.

[RDGF16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. IEEE conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[RF17] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7263–7271, 2017.

[RG09] Miquel Ramírez and Hector Geffner. Plan recognition as planning. In *Proc. IJCAI-09*, pages 1778–1783, 2009.

[Rin17] Jussi Rintanen. Schematic invariants by reduction to ground invariants. In *AAAI*, 2017.

[RW10] Sylvia Richter and Matthias Westphal. The LAMA planner: Guiding cost-based anytime planning with landmarks. *Journal of Artificial Intelligence Research*, 39(1):127–177, 2010.

[SA77] Roger C Schank and Robert P Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Lawrence Earlbaum, 1977.

[SAJJ16] Javier Segovia-Aguas, Sergio Jiménez, and Anders Jonsson. Hierarchical finite state controllers for generalized planning. In *Proc. IJCAI*, 2016.

[SB98] Richard Sutton and Andrew Barto. *Introduction to Reinforcement Learning.* MIT Press, 1998.

[SG16] Luciano Serafini and Artur d'Avila Garcez. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. *arXiv preprint arXiv:1606.04422*, 2016.

[SHM+16] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.

[SHS+18] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.

[SIZ08] Siddharth Srivastava, Neil Immerman, and Shlomo Zilberstein. Learning generalized plans using abstract counting. In *AAAI*, volume 8, pages 991–997, 2008.

[SIZ11] Siddharth Srivastava, Neil Immerman, and Shlomo Zilberstein. A new representation and associated algorithms for generalized planning. *Artificial Intelligence*, 175(2):615–647, 2011.

[SPLC16] Zhao Song, Ronald E Parr, Xuejun Liao, and Lawrence Carin. Linear feature encoding for reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 4224–4232, 2016.

[SPS99] Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

[SRB+17] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. In *Advances in neural information processing systems*, pages 4967–4976, 2017.

[SRBS16] Martin EP Seligman, Peter Railton, Roy F Baumeister, and Chandra Sripada. *Homo prospectus*. Oxford University Press, 2016.

[SSS+17] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.

[SYZ+18] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, 2018.

[TBF+18] Valentin Thomas, Emmanuel Bengio, William Fedus, Jules Pondard, Philippe Beaudoin, Hugo Larochelle, Joelle Pineau, Doina Precup, and Yoshua Bengio. Disentangling the independently controllable factors of variation by interacting with the world. *arXiv preprint arXiv:1802.09484*, 2018.

[TS11] Matthew E Taylor and Peter Stone. An introduction to intertask transfer for reinforcement learning. *AI Magazine*, 32(1):15, 2011.

[TTTX18] Sam Toyer, Felipe Trevizan, Sylvie Thiébaux, and Lexing Xie. Action schema networks: Generalised policies with deep learning. In *AAAI*, 2018.

[XZF+18] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Broeck. A semantic loss function for deep learning with symbolic knowledge. In *ICML*, pages 5498–5507, 2018.

[YLWA05] Håkan LS Younes, Michael L Littman, David Weissman, and John Asmuth. The first probabilistic track of the international planning competition. *Journal of Artificial Intelligence Research*, 24:851–887, 2005.

[YWJ07] Qiang Yang, Kangheng Wu, and Yunfei Jiang. Learning action models from plan examples using weighted max-sat. *Artificial Intelligence*, 171(2-3):107–143, 2007.

[ZK15] Sergey Zagoruyko and Nikos Komodakis. Learning to compare image patches via convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4353–4361, 2015.