Hierarchical principles of embodied reinforcement learning: A review

Manfred Eppe1,*, Christian Gumbsch2,3, Matthias Kerzel1, Phuong D.H. Nguyen1, Martin V. Butz2, and Stefan Wermter1

1Universität Hamburg, Germany
2Universität Tübingen, Germany
3Max Planck Institute for Intelligent Systems, Tübingen, Germany
*eppe@informatik.uni-hamburg.de

ABSTRACT

Cognitive Psychology and related disciplines have identified several critical mechanisms that enable intelligent biological agents to learn to solve complex problems. There exists pressing evidence that the cognitive mechanisms that enable problem-solving skills in these species build on hierarchical mental representations. Among the most promising computational approaches to provide comparable learning-based problem-solving abilities for artificial agents and robots is hierarchical reinforcement learning. However, so far the existing computational approaches have not been able to equip artificial agents with problem-solving abilities that are comparable to intelligent animals, including human and non-human primates, crows, or octopuses. Here, we first survey the literature in Cognitive Psychology, and related disciplines, and find that many important mental mechanisms involve compositional abstraction, curiosity, and forward models. We then relate these insights with contemporary hierarchical reinforcement learning methods, and identify the key machine intelligence approaches that realise these mechanisms. As our main result, we show that all important cognitive mechanisms have been implemented independently in isolated computational architectures, and there is simply a lack of approaches that integrate them appropriately. We expect our results to guide the development of more sophisticated cognitively inspired hierarchical methods, so that future artificial agents achieve a problem-solving performance on the level of intelligent animals.

Humans and several other higher level intelligent animal species have the ability to break down complex unknown problems into hierarchies of simpler previously learned sub-problems. This hierarchical approach allows them to solve previously unseen problems in a zero-shot manner, i.e., without any trial and error. For example, Fig. 1 depicts how a New Caledonian crow solves a non-trivial food-access puzzle that consists of three causal steps: It first picks a stick, then uses the stick to access a stone, and then uses the stone to activate a mechanism that releases food. There exist numerous analogous experiments that attest similar capabilities to primates, octopuses, and, of course, humans. This raises the question of how we can equip intelligent artificial agents and robots with similar problem-solving abilities. To answer this question, the involved computational mechanisms and algorithmic implementation options need to be identified.

A very general computational framework for learning-based problem-solving is reinforcement learning (RL). Several studies suggest that RL is in many aspects biologically and cognitively plausible. Existing RL-based methods are to some extent able to perform zero-shot problem-solving and transfer learning. However, this is currently only possible for minor variations of the same or a similar task, or in simple synthetic domains, such as a 2D gridworld. A continuous-space problem-solving behaviour that is comparable with the crow’s behaviour of Fig. 1 has not yet been realised with any artificial system, based on RL or other approaches.

Research in human and animal cognition strongly suggests that problem-solving and learning is hierarchical. We hypothesise that one reason why current machine learning systems fail is that the existing approaches underestimate the power of learning hierarchical abstractions. At the time of writing this article, we performed a comprehensive meta-search over RL review articles from 2015 to 2020 using the Microsoft Academic search engine. Our meta-survey has yielded the following results: The most cited review article on RL since 2015 dedicates 1/6th of a page out of 16 pages to hierarchical approaches. In the second most cited article, hierarchical RL is not considered at all, i.e. the word stem ‘hierarchic’ does not appear anywhere in the text. The third most cited review article dedicates 2/3rd of a page out of 85 total pages to hierarchical RL. From 37 RL reviews and surveys published since 2015, only two contain the word stem “hierarchic” in their abstract. The second edition of the popular RL book by Sutton and Barto only mentions hierarchies in the twenty-year-old options framework on pages in the final book chapter and briefly discusses automatic abstraction a few lines later in that chapter. It appears that researchers are struggling with automatically learning RL-suitable hierarchical structures.

In this article, we address this gap by illuminating potential reasons for this struggle and by providing pointers to solutions. We show that most computational hierarchical RL approaches are model-free, whereas the results from our survey of biological cognitive models suggest that suitable predictive forward models are needed. Furthermore, we exhibit that existing hierarchical RL methods hardly address state abstraction and compositional representations. However, we also show that there exist exceptions where isolated relevant cognitive mechanisms including forward models and compositional abstraction have already been implemented, but not in combination with other important mechanisms. As a result, we conclude that the AI community already has most of the tools for building more intelligent agents at hand; but it currently lacks holistic efforts to integrate them appropriately.

Neurocognitive foundations

The problem-solving abilities of a biological agent depend on cognitive key mechanisms that include abstraction, intrinsic motivation, and mental simulation. Fig. 2 shows how these mechanisms depend on forward and inverse models as neuro-functional prerequisites. Together, cognitive abilities that we deem crucial for higher level intelligence become enabled, including few-shot problem solving and...
transfer learning. In the following, we start with a characterization of these crucial cognitive abilities and then exhibit some of the key mechanisms and prerequisites needed to enable them.

**Crucial cognitive abilities**

By cognitive abilities for problem-solving we refer to phenomenologically observable traits and properties of biological agents. As the most remarkable ability, we consider few-shot problem-solving.

**Few-shot problem-solving** is the ability to solve unknown problems with few (≤ 10) trials. For simple problems, few-shot problem-solving is trivial. For example, catching a ball is usually a purely reactive behaviour that builds on a direct neural mechanism to map the observation of the flying ball to appropriate motor commands. Such problems are solvable with current computational RL methods. However, there are also more difficult problems that require, e.g., using tools in new contexts. For example, consider the aforementioned problem-solving example of the crow (see Fig. 1). The crow knows how to use a stick as a tool without any further training, because it has previously used a stick for related problems. In our conceptual framework (see Fig. 2) we identify two abilities that are most central to perform such problem-solving, namely transfer learning and planning.

**Transfer learning** allows biological agents to perform few-shot problem-solving by transferring the solution to a previously solved task to novel, previously unseen tasks. This significantly reduces and sometimes eliminates the number of trials to solve a problem.

Cognitive theories consider analogical reasoning as a critical cornerstone of human cognition. For example, conceptual metaphor theory understands metaphors as mappings between analogous concepts in different domains. In addition to evidence from linguistics, there is also significant evidence from education theory and research on mechanical reasoning that analogical reasoning improves problem-solving: Education theory suggests that human transfer learning improves when explicitly trained to identify analogies between different problems. The Psychology of mechanical
inference suggests that analogical knowledge about the dynamic properties of mechanical objects is often transferred between domains. For example, knowledge about what happens if somebody jumps into a pool can be transferred to other mechanical problems that involve over-flooding water containers.

Goal-directed planning. The behaviour of biological agents is traditionally divided into two categories: Stimulus-driven, habituated behaviour and goal-directed, planned behaviour. Initially, stimulus-response theories dominated the field, suggesting that most behaviour was learned by forming associations that were previously reinforced. Edward Tolman was one of the main critiques of stimulus-response theories. He showed that rats can find rewards in a maze faster when they have visited the maze before, even if their previous visit had not been rewarded. The results suggest that the rats form a representation, or cognitive map, of the maze, which enables them to plan or mentally simulate their behaviour once a reward was detected. Habituation of behaviour, which is comparable to model-free policy optimization in reinforcement learning, enables an agent to learn one particular skill in a controlled environment. However, complex problem-solving in a new or changing environment requires some form of prospective planning. Over the last decades, more research has focused on understanding the mechanisms and representations involved in goal-directed planning. It is now generally agreed upon that, during planning, humans predict the effect of their actions, compare them to the desired effects, and, if required, refine their course of actions. This is deemed to be a hierarchical process where the effects of actions on different levels of abstraction are considered. Thereby, humans tend to first plan high-level actions before considering actions at a finer granularity. Hierarchical planning can dramatically decrease the computational cost involved in planning. Additionally, hierarchical abstractions enable automatizing originally planned behaviour, thus further alleviating computational cost.

Cognitive mechanisms for transfer learning and planning

Transfer learning and planning rely on shared and partially overlapping mechanisms and principles. Here, we focus on three mechanisms and principles that we identify as most critical, namely sensorimotor abstraction, intrinsic motivation, and mental simulation.

Sensorimotor abstraction.

According to the widely accepted embodiment theory, abstract mental concepts are inferred from our bodily-grounded sensorimotor environmental interaction experiences. Cognitive theories often distinguish between action abstraction and state abstraction. Action abstractions refer to a temporally extended sequence of primitive actions, that are often referred to as options or movement primitives. For example, transporting an object can be seen as an action abstraction since it is composed of a sequence of more primitive actions, such as reaching, lifting, and transporting (see Fig. 3). The elementary actions that compose an action abstraction are typically rather loosely defined through their effects. For example, we can tell that a robot arm we have never seen before is grasping an object, even though the anatomy, the unfolding control algorithms, and the involved motor commands may largely vary from our human ones. State abstractions refer to encoding certain parts of the environment while abstracting away irrelevant details. A simple example of a state abstraction is a code that specifies whether a certain type of object is present in a scene. State abstractions can be hierarchically structured: A single object can be subdivided into its parts to form a partonomy, i.e., a hierarchical organization of its parts and their spatial relationship (see Fig. 3). Additionally, objects can be organized taxonomically, reflecting a hierarchical relationship that involves particular concept hierarchies. For example, an affordance-oriented taxonomy of beverage containers could classify wine glasses and beer mugs both as drinking vessels, while a teapot would be classified as a pot instead. Meanwhile, beer mugs, wine glasses, and teapots are all graspable containers.

Conceptual abstractions and cross-references to other conceptual abstractions within imply a representational key property: compositionality. Formal compositionality principles state that an expression or representation is compositional if it consists of subexpressions, and if there exist rules that determine how to combine the subexpressions. The Language of Thought theory transfers the compositionality principle from language to abstract mental representations, claiming that thought must also be compositional. For example, humans are easily able to imagine previously unseen situations and objects by composing their known constituents, such as the golden pavements and ruby walls imagined by Scottish philosopher David Humes in A Treatise of Human Nature.

On top of that, embodied world knowledge further constrains the formal semantics-based options to combined sub-expressions. In the remainder, we will refer to this type of compositionality by common sense compositionality. This is, compositional rules that are grounded in and flexibly inferred by our intuitive knowledge about the world, including other agents (humans, animals, robots, machines, etc.). This common sense compositionality essentially enforces that our minds attempt to form plausible environmental scenarios and events that unfold within given conceptual information about a particular situation. Consider the processes that unfold when understanding a sentence like “He sneezed the napkin off the table”. Most people have probably never heard or used this sentence before, but everyone with sufficient world and English grammar knowledge can correctly interpret the sentence, imagining a scene wherein the described event unfolds. Common sense compositionality makes abstractions applicable in meaningful manners, by constraining the filling in of an abstract component (e.g., the target of a grasp) towards applicable objects (e.g., a graspable object such as a teapot), as depicted in Fig. 3. Note the awkwardness and indeed the attempt of the brain to resolve it—when disobeying common sense compositionality, such as when stating that “The dax fits into the air” or “grasping a cloud”.

Common sense compositionality indeed may be considered a hallmark and cornerstone of human creativity, as has been (indirectly) suggested by numerous influential cognitivists. In the neurosciences, recent neuroimaging techniques indeed revealed that the compositionality of mental representations can actually be observed in neural codes. For example, Haynes et al. and Reverberi et al. show that the neural codes in the ventrolateral prefrontal cortex representing compound behavioural rules can be decomposed into neural codes that encode their constituent rules. Similar results have been obtained for neural codes that represent food compositions.

From an algorithmic perspective, common sense compositionality yields benefits for analogical reasoning and planning. It simplifies the identification of analogies because it enables compositional morphisms between representations. This advantage is well-known in cognitive theories of creativity, e.g., concept blending, where
the search for analogies is a computational bottleneck. Similarly, compositional state- and action representations for goal-directed planning lead to a lower computational complexity, because action and object types can be flexibly combined as long as critical properties match. Along these lines, action grammars have been proposed, which systematically form an effect-oriented taxonomy of tool-object interactions. As a result, common sense compositionality enables the utilisation of objects as tools in innovative manners—when, for example, utilising a stone as a hammer. Moreover, it enables drawing analogies across domains—when, for example, talking about “holding onto a thought” for later utilisation.

Common sense compositionality thus seems to be key for truly intelligent artificial systems. We propose that apart from the addition of suitable inductive learning biases, such as event-segmentation tendencies, suitably targeted intrinsic motivation and the ability to play out mental simulations are of critical importance.

**Intrinsic motivation** affects goal-directed reasoning and planning because it sets intrinsically motivated goals that an agent may aim to achieve. The most basic behaviour of biological agents purely strives for satisfying homeostatic needs, such as hunger or thirst. However, advanced behaviour, such as exploration and play, seems to be partially decoupled from the primary biological needs of an animal. From a cognitive development perspective, the term intrinsic motivation was coined to describe the “inherent tendency to seek out novelty, [...] to explore and to learn”.[12] Here, intrinsic is used in opposition to extrinsic values that are directly associated with the satisfaction of needs or the avoidance of undesired outcomes.[13] Intrinsic motivations induce general exploratory behaviour, curiosity, and playfulness. Simple exploratory behaviour can even be found in worms, insects, and crustaceans and may be elicited by rather simple tendencies to wander around while being satiated or an in-bored tendency to evade overcrowded areas. Curiosity refers to an epistemic drive that causes information gain-oriented exploratory behaviour.[14] Combined with developing common sense compositionality, this drive can elicit hypothesis testing behaviour even by means of analogy making. The closely related playfulness is only exhibited in certain intelligent, mostly mammalian or avian species, such as dogs and corvids, where scenarios and events within are played out in a hypothetical manner and novel interactions are tried-out in a skill-improvement-oriented manner.

**Mental simulation** meanwhile, enables biological agents to reflect on and to anticipate the dynamics of its environment on various representational levels in the past, present, and future, and even fully hypothetically. Actual motor-imagery has been shown to enable athletes to improve the execution of challenging body movements (e.g. a backflip), significantly reducing the number of actually executed trials.[15] On the reasoning side, consider human engineers who rely on mental simulation when developing mechanical systems, such as pulleys or gears. The simulation involves visual imaginations, but also includes non-visual properties like friction and force.[16] Mental simulation also takes place on higher conceptual, compositional, and causal reasoning levels. For example, Kahnemann and Tversky, and Wells and Gavanski, report how mental simulation improves the planning of future behaviour on a causal conceptual level, such as when planning a trip.

**Forward-inverse models as functional prerequisites**

Sensorimotor abstraction, intrinsic motivation, and mental simulation are complex mechanisms that demand a suitable neuro-functional basis. We propose that this essential basis is constituted by forward and inverse models.[17] Forward models predict how an action affects the world, while inverse models determine the actions that need to be executed to achieve a goal or to obtain a reward. Note that inverse models may be implicitly inferred from the available forward models, but more compact inverse models, which are closely related to habits and motion primitives, certainly foster the proficiency of particular behavioural skills further.[18,19,20,21,22]

**Forward-inverse models for mental simulation.** To perform mental simulation, an agent needs a forward model to simulate how the dynamics of the environment behave, possibly conditioned on the agent’s actions. However, even when equipped with a well-predicting forward model, the consideration of all possible actions and action sequences quickly becomes computationally intractable. Hence, action-selection requires more direct mappings from state and desired inner mental state (including an intention or a goal) to potential actions. This is accomplished by inverse models. In RL, inverse models are represented as behaviour policies. Selecting actions can happen in a reflex-like manner on the motor level but it can also serve as a heuristic to generate candidate actions for accomplishing a current goal. For example, when designing a new mechanical system, engineers have an intuition about how to put certain mechanical parts together, depending on the goal function they want to achieve with the system. Based on their intuition, they mentally simulate and verify the functioning of the mechanical system.[23]

**Forward and inverse models for intrinsic motivation.** Forward and inverse models are also needed to trigger behaviour that is driven by intrinsic motivation. For example, prediction error minimisation has been demonstrated to be a useful additional, intrinsic reward signal to model curiosity, inversely triggering behaviour and behavioural routines that are believed to cause particular environmental changes that are, in turn, believed to possibly result in knowledge gain.[24,25,26,27,28] Along similar lines, Friston et al. propose that intrinsically motivated behaviour emerges when applying active inference.[29,30] Active inference describes the process of inferring actions to minimize expected free energy, which includes approximations of anticipated surprise.[31] Free energy is decomposed into various sources of uncertainty. One part is uncertainty about future states or outcomes given a certain sequence of actions.[32] The agent strives to reduce this uncertainty, and with it overall expected free energy, by activating the actions that cause the uncertainty. Hence, active inference can lead to epistemic exploration, where an agent acts out different behaviour to learn about its consequences.[33] Here, the forward model is required to predict future states given an action, and also to estimate the uncertainty of the prediction.

**Forward and inverse models for abstraction and grounding.** Over the last decade, various theories including predictive coding,[34] the Free Energy Principle,[35] and the Bayesian Brain hypothesis,[36] have viewed the brain as a generative machine, which constantly attempts to match incoming sensory signals with its probabilistic predictions.[37] Within these frameworks, prediction takes place on multiple processing hierarchies that interact bidirectionally: Top-down information per processing stage provides additional information to predict sensory signals, while bottom-up error information is used to correct the top-down predictions.[38] Meanwhile, future states are predicted in a probabilistic manner.[39] On lower processing levels, actual incoming sensorimotor signals are predicted, while on higher levels more abstract, conceptual and compositional predictive encodings emerge.[40] In this way, rather complex state abstractions, such as the aforementioned container-concept, can form. As a result, high-level predictions, e.g., of an object being within a container, enable predictions on lower levels. For example, consider the prediction of sensory information about how an object’s position will change over time while it is occluded.

Event Segmentation Theory (EST) makes the role of forward predictions even more explicit: EST is concerned with why humans seem to perceive the continuous sensory activity in terms of discrete, conceptual events. The theory suggests that these segmentations mirror the internal representation of the unfolding experience. Internal representations of events, or event models, are sensorimotor abstractions that encode entities with certain properties and their
A significant fraction of the existing near transfer approaches build on the automatic discovery of transferable macro-actions (plans) to solve problems in discrete 2D-environments, such as navigation skills acquired in a discrete-space grid maze environment to a continuous-space ant maze environment. These approaches underpin that planning is especially useful for high-level inference in discrete state-action spaces.

**Mechanisms behind transfer learning and planning**

Our summary of the cognitive principles behind transfer learning and planning reveals three important mechanisms that are critical for the learning and problem-solving capabilities of biological agents, namely compositional sensorimotor abstraction, intrinsic motivation, and mental simulation.

**Compositional sensorimotor abstraction and grounding.** The temporal abstraction of actions is, by definition, an inherent property of hierarchical reinforcement learning as it allows to break down complex problems into a temporal hierarchical structure of simpler problems. Another dimension of abstraction is representational abstraction. Hierarchical reinforcement learning approaches also involve representational action abstraction. One can distinguish the existing approaches into two different types of representational action abstraction. The probably most influential method for action abstraction builds on behaviour primitives that are successful. In this sense, planning differs from model-based reinforcement learning, which usually refers to training a policy by simulating the actions with an internal model of the domain dynamics. Planning enables one-shot problem-solving because the searching does not involve the physical execution of actions. The agent only executes the actions if the mental simulation verifies that the actions are successful. In this sense, planning differs from model-based reinforcement learning, which usually refers to training a policy by simulating actions using a predictive forward model.

Hierarchical planning is a well-known paradigm in classical AI, but approaches that integrate planning with hierarchical reinforcement learning are rare. Some approaches integrate action planning with reinforcement learning by using an action planner for high-level decision-making and a reinforcement learner for low-level motor control. These approaches underpin that planning is especially useful for high-level inference in discrete state-action spaces.

**Computational realizations**

The abilities, mechanisms and prerequisites of computational hierarchical reinforcement learning systems (cf. Fig. 4) are less sophisticated and integrated than those of biological agents. However, there are promising novel developments to potentially overcome the existing limitations. To identify the potential of the existing computational approaches, we provide an overview of the current state of the art on hierarchical reinforcement learning in Tab. 1.

**Transfer learning and planning for few-shot abilities**

Our survey of the neurocognitive foundations indicates that two foundational cognitive abilities for few-shot problem-solving are transfer learning and planning.

**Transfer learning** denotes the re-use of previously learned skills in new contexts and domains, where near transfer learning denotes transfer between similar contexts and domains between source and target tasks, while far transfer considers stronger dissimilarities. A significant fraction of the existing near transfer approaches build on learning re-usable low-level skills, which are referred to as behavioural primitives, options, or sub-policies. For example, Li et al. and Frans et al. present sub-policy-based hierarchical extensions to Proximal Policy Optimization (PPO). Their approaches enable the transfer of sub-policies to novel tasks within the same domain. Heess et al. use Generalized Advantage Estimation (GAE) to implement similar transferable sub-policies. However, the authors also consider transfer learning towards different types of tasks. For example, they transfer behaviour from a locomotion task to a more complex soccer task. Eysenbach et al. and Sharma et al. introduce diversity functions to learn diverse re-usable sub-policies that are independent of external rewards.

Tessler et al. focus on the transfer of skills in lifelong learning, and Wu et al. propose a model-based approach, where only the forward models are transferred to new tasks, but not the policies. Vezhnevets et al. build on the automatic discovery of transferable macro-actions (plans) to solve problems in discrete 2D-environments, while Jiang et al. use natural language action representations to perform near transfer learning. Qureshi et al. build on re-usable low-level policies for goal-conditioned hierarchical reinforcement learning.

Research has not only considered transfer learning between different tasks but also between different robot and agent morphologies. We classify these approaches as far transfer because the entire sensorimotor apparatus changes, which places the agent in a dissimilar context. Furthermore, the methods that feature such cross-morphological transfer also perform far transfer learning across different application domains. For example, Frans et al. transfer navigation skills acquired in a discrete-space grid maze environment to a continuous-space ant maze environment.

Fig. 4 | A general hierarchical problem-solving architecture with a separate policy $\pi_i$ for each layer of abstraction. Most existing computational HRL approaches focus on two layers, where the high-level action $a'$ is either a behavioural primitive (an option or sub-policy), or a subgoal. Only the low-level actions $a^0$ are motor commands that affect the environment directly.
Reach pot
Move pot
Pour tea
Initial state
Pot grasped
Tea in cup!
Pot above cup
Tea in cup!
Initial state

Fig. 5 | Action abstraction through behaviour primitives (a) vs subgoals (b). With behaviour primitives, the agent determines the path to the final goal by selecting a sequence of high-level actions, but without specifying explicitly to which intermediate state each action leads. With subgoals, the agent determines the intermediate states as subgoals that lead to the final goal, but without specifying the actions to achieve the subgoals.

ably less research on state abstraction in hierarchical reinforcement learning. Cognitive state abstractions range from the mere preprocessing of sensor data, e.g. in the primary visual cortex, to the problem-driven grounding of signals in abstract compositional concept representations. Counterparts for some of these methods can also be found in computational architectures. For example, most current reinforcement learning approaches that process visual input use convolutional neural networks to preprocess the visual sensor data. A problem with simple preprocessing is that it does not appreciate that different levels of inference require different levels of abstraction: For example, to transport an object from one location to another with a gripper, only the low-level motor control layer of a hierarchical approach needs to know the precise shape and weight of the object. A high-level planning or inference layer only needs to know abstract Boolean facts, such as whether the object is light enough to be carried.

Therefore, we consider only those approaches as representational abstraction methods that involve layer-wise abstraction. Layer-wise state abstraction has been tackled, but most existing hierarchical reinforcement learning approaches perform the abstraction using manually defined abstraction functions.64,74,82,126
a few exceptions where state abstractions are derived automatically in hierarchical reinforcement learning architectures\textsuperscript{138,139}. For instance, through clustering\textsuperscript{134}, with feature representations\textsuperscript{106} or by factorisation\textsuperscript{130}. Interestingly, non-hierarchical model-based reinforcement learning offers promising alternative prediction-based methods for state abstraction\textsuperscript{2,92}, which show parallels to cognitive prediction-based abstraction theories. However, these have not yet been applied in a hierarchical architecture.

As implied from the cognitive science side, a key-property of representations of states and actions is compositionality. For instance, a symbolic compositional action representation $\text{grasp(\text{glass})}$ allows for modulating the action $\text{grasp}$ with the object $\text{glass}$. Compositionality is not limited to symbolic expressions, but also applicable to distributed numerical expressions, such as vector representations. For example, a vector $v_1$ is a compositional representation if it is composed of other vectors, e.g., $v_2$ and $v_3$, and if there is a vector operation $\circ$ that implements interpretation rules, e.g., $v_1 = v_2 \circ v_3$.

There is significant cognitive evidence that compositionality improves transfer learning\textsuperscript{39}. This evidence is computationally plausible when considering that transfer learning relies on exploiting analogies between problems. The analogies between two or more problems in goal-conditioned reinforcement learning are defined by a multidimensional mapping between the initial state space, the goal space, and the action space of these problems. For example, given $n_a = 4$ action types (e.g., “grasp”, “push”, “move”, “release”) and $n_o = 4$ object types (e.g., “glass”, “cup”, “tea pot”, “spoon”), a non-compositional action-representations requires one distinct symbol for each action-object combination, resulting in $n_a \times n_o / 2$ possible action mappings. In contrast, an analogy mapping with compositional actions would require to search over possible mappings between action types and, separately, over mappings between objects. Hence, the size of the search space is $n_a / 2 + n_o / 2$. Evidence that a lower number of possible analogy mappings improves transferability is also provided by other cognitively inspired computational methods, such as concept blending\textsuperscript{28}.

The work by Jiang et al.\textsuperscript{64} provides further empirical evidence that compositional representations improve learning transfer. The authors use natural language as an inherently compositional representation to describe high-level actions in hierarchical reinforcement learning. In an ablation study, they find that the compositional natural language representation of actions improves transfer learning performance compared to non-compositional representations\textsuperscript{447,106}. Few other researchers use compositional representations in hierarchical reinforcement learning. Sax et al.\textsuperscript{103} compose high-level actions from concurrent linearly solvable Markov decision processes (LMDPs) to guarantee optimal compositionality. Their method successfully executes compositional policies that it has never executed before, effectively performing zero-shot problem-solving. Zero-shot problem-solving has also been demonstrated by other related research that features compositional action representations, but that does not draw an explicit link between compositionality and zero-shot problem-solving\textsuperscript{81,88,115}. Few symbolic compositional state representations exist, but these rely on manually defined abstraction functions\textsuperscript{29,69,81,126} or they are very general and feature compositionality only as an optional property\textsuperscript{106}. There also exist hierarchical reinforcement learning approaches where sub-symbolic compositional representations are learned\textsuperscript{9,24,97,106,130}. The compositionality of these representations is rather implicit and, with one exception\textsuperscript{29}, has not been investigated in the context of transfer learning. The exceptional approach by Qureshi et al.\textsuperscript{77} considers composable low-level policies and shows that compositionality significantly improves transfer between similar environments.

**Intrinsic motivation** is a useful method to stabilise reinforcement learning by supplementing sparse external rewards. It is also commonly used to incentivize exploration. Reinforcement learning typically models intrinsic motivation through intrinsic rewards. The most common method of hierarchical reinforcement learning to generate intrinsic rewards is by providing intrinsic rewards when subgoals or subtasks are achieved\textsuperscript{9,15,29,30,82,69,81,88,97,100,102,103,104,113,138,139}. Other approaches provide intrinsic motivation to identify a collection of behavioural primitives with a high diversity\textsuperscript{9,30,83} and predictability\textsuperscript{115}. This includes also the identification of primitives that are suited for re-composing high-level tasks\textsuperscript{112}.

Another prominent intrinsic reward model that is commonly used in non-hierarchical reinforcement learning is based on surprise and curiosity\textsuperscript{9,15,29,90,101,107,109}. In these approaches, surprise is usually modelled as a function of the prediction error of a forward model, and curiosity is realised by providing intrinsic rewards if the agent is surprised. However, there is only little work on modelling surprise and curiosity in hierarchical reinforcement learning. Only Blaes et al.\textsuperscript{9}, Colas et al.\textsuperscript{19}, and Röder et al.\textsuperscript{102} use surprise in a hierarchical setting, showing that hierarchical curiosity leads to a significant improvement of the learning performance. These approaches train a high-level layer to determine explorative subgoals in hierarchical reinforcement learning. An alternative method to model curiosity is to perform hypothesis-testing for option discovery\textsuperscript{16}.

**Mental simulation** is a mechanism that allows an agent to anticipate the effects of its own and other actions. Therefore, it is a core mechanism to equip an intelligent agent with the ability to plan ahead. Cognitive theories about mental simulation involve predictive coding\textsuperscript{29} and mental motor imagery\textsuperscript{49}, while computational approaches to mental simulation involve model-based reinforcement learning\textsuperscript{113}, action planning\textsuperscript{96}, or a combination of both\textsuperscript{39,51}. However, even though there is significant cognitive evidence that mental simulation happens on multiple representational layers\textsuperscript{13,15}, there is a lack of approaches that use hierarchical mental simulation in hierarchical reinforcement learning. Only a few approaches that integrate planning with reinforcement learning build on mental simulation\textsuperscript{9,81,82,113,138,139}, though the mental simulation mechanisms of these models are only implemented on the high-level planning layer. An exception is the work by Wu et al.\textsuperscript{136}, who use mental simulation on the low-level layer to determine the sub-policies to be executed. Another exception is presented by Li et al.\textsuperscript{79} who perform mental simulation for planning on multiple task layers.

There exist several non-hierarchical model-based reinforcement learning methods\textsuperscript{49,52,118} that function akin to mental motor imagery: In these approaches, the policy is trained by mentally simulating action execution instead of executing the actions in the real world. However, even though mental simulation is deemed to be a hierarchical process, we are not aware of any approach that performs hierarchical model-based reinforcement learning in the classical sense, where the policy is trained on the developing forward model.

**Prerequisites for sensorimotor abstraction, intrinsic motivation and mental simulation**

Reinforcement learning builds on policies that select actions based on the current observation and a goal or a reward function. Policies can be modelled directly, derived from value functions or combined with value functions. In all cases, a policy is an inverse model that predicts the actions to be executed in the current state to maximise reward or to achieve a goal state. In contrast, a forward model predicts a future world state based on the current state and a course of actions. Both kinds of models are critical prerequisites for the mechanisms that enable transfer learning, planning, and ultimately few-shot problem-solving. Our review shows that the vast majority of hierarchical reinforcement learning methods use inverse models for both the high-level and low-level layers. However, some approaches exist that use an inverse model only for the low-level layer\textsuperscript{26,81,113,138,139}. These frameworks use a planning mechanism, driven by a forward model, to perform the high-level decision-making. Our review demonstrates that a forward model is required for several additional mechanisms that are necessary or at least highly
beneficial for transfer learning, planning, and few-shot problem-solving. Some non-hierarchical approaches use a forward model to perform sensorimotor abstraction\cite{29, 91}. They achieve this with a self-supervised process where forward predictions are learned in a latent abstract space. However, we are not aware of any hierarchical method that exploits this mechanism.

A forward model is also required to model curiosity as an intrinsic motivation mechanism. Cognitive science suggests that curiosity is one of the most important drives of human development, and it has been demonstrated to alleviate the sparse rewards problem in reinforcement learning\cite{91, 103}. Curiosity is commonly modelled as a mechanism that rewards an agent for being surprised; surprise is formally a function of the error between the predicted dynamics of a system and the actual dynamics\cite{39}. To perform these predictions, a forward model is required. However, with few exceptions\cite{19, 103}, there is a lack of approaches that use a hierarchical forward model for generating hierarchical curiosity.

**Results**

Tab. 1 shows the results of our review on computational hierarchical reinforcement learning approaches, in alignment with our structuring of cognitive prerequisites and mechanisms for the problem-solving abilities of intelligent biological agents (see Fig. 2). We summarise our main results as follows:

**Most current few-shot problem-solving methods build on transfer learning but they do not leverage planning.** This result is interesting because any approach with a forward model could straightforwardly be extended to also consider planning\cite{10}, and planning can be leveraged for few-shot problem-solving\cite{18}. Therefore, current model-based approaches do not exploit their full potential.

**Current methods do not exploit the full potential of compositional abstract representations.** All hierarchical reinforcement learning methods inherently perform representational abstraction, e.g., through options or subgoals, but only a few methods consider compositional abstraction. Exceptions include natural language-based representations of actions\cite{64} or symbolic logic-based compositional representations\cite{88}. None of these approaches ground these abstractions tightly to sensorimotor experiences.

**State abstraction is significantly less researched than action abstraction.** Recent hierarchical actor-critic approaches\cite{76, 103} use the same state representation for each layer in the hierarchy, without performing any abstraction. There exist a few approaches that perform state abstraction\cite{329} and also some that involve compositional state representations\cite{130}. However, most of these build on hand-crafted abstraction functions that are not learned\cite{29}.

**Curiosity and diversity, as intrinsic motivation methods, are underrepresented.** When assessing the intrinsic motivation mechanisms of the existing approaches, we distinguish between the three most frequently used kinds of intrinsic motivation: curiosity, diversity, and subgoal discovery. Tab. 1 indicates that curiosity and diversity are underrepresented intrinsic motivation methods, even though there is significant cognitive evidence that these mechanisms are critical for intelligent behaviour in humans\cite{89}.

Finally, **there is a lack of hierarchical mental simulation methods.** While numerous non-hierarchical methods show how mental simulation can improve sample efficiency\cite{23, 49}, our summary shows that reinforcement learning with hierarchical mental simulation is yet to be explored. As a result, inter-dependencies between sub-tasks are hard to detect.

**Conclusion**

Our review provides an overview of the cognitive foundations of hierarchical problem-solving and how these are implemented in current hierarchical reinforcement learning architectures. Herein, we focus on few-shot problem-solving – the ability to learn to solve causally non-trivial problems with as few trials as possible or, ideally, without any trials at all (zero-shot problem-solving). As a benchmark example, we refer to recent results from animal cognition\cite{47}, demonstrating how a crow develops mental representations of tools to solve a previously unseen food-access problem at the first trial (see Fig. 1). As our main research question, we ask for the computational prerequisites and mechanisms to enable problem-solving for artificial computational agents that are on the level of intelligent animals.

We provide two main contributions to address this question. First, we perform a literature survey and a top-down structuring of cognitive abilities, mechanisms, and prerequisites for few-shot problem-solving (see Fig. 2). Second, we perform a comprehensive review of recent hierarchical reinforcement learning architectures to identify whether and how the existing approaches implement the prerequisites, mechanisms, and abilities towards few-shot problem-solving (see Tab. 1). Herein, we identify five major shortcomings in the state of the art of the computational approaches. These shortcomings are mostly based on the lack of hierarchical forward models and compositional abstractions.

Nonetheless, we were able to identify several methods that address these gaps in the state of the art in isolation. Not seeing any major reason why these approaches could not be integrated, we suggest that most of the tools to realise higher levels of intelligent behaviour in artificial agents have already been investigated. The key is to combine them effectively. This demonstrates significant potential to develop intelligent agents that can learn to solve problems on the level of intelligent biological agents.

**Acknowledgements**

Manfred Eppe, Phuong Nguyen and Stefan Wermter acknowledge funding by the German Research Foundation (DFG) through the projects IDEAS and LeCAREbot. Stefan Wermter and Matthias Kerzel acknowledge funding by the DFG through the transregio project TTR 169 Cross-modal Learning (CML). Christian Gumbsch and Martin Butz acknowledge funding by the DFG project number BU 1335/11-1 in the framework of the SPP program “The Active Self” (SPP 2134). Furthermore, Martin Butz acknowledges support from a Feodor-Lynen stipend of the Humboldt Foundation as well as from the DFG, project number 198647426, “Research Training Group 1808: Ambiguity - Production and Perception” and from the DFG-funded Cluster of Excellence “Machine Learning – New Perspectives for Science” (EXC 2064/1, project number 390727645). The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Christian Gumbsch.

**References**

1. Riad Akrour, Filipe Veiga, Jan Peters, and Gerhard Neumann. Regularizing reinforcement learning with state abstraction. In Proc. of International Conference on Intelligent Robots and Systems (IROS), 534–539 (2018).
2. Kai Arulkumaran, Nat Dilkushanakul, Murray Stratahan, and Anil Anthony Bharath. Classifying options for deep reinforcement learning. In Proc. of IJCAI Workshop on Deep Reinforcement Learning: Frontiers and Challenges. http://arxiv.org/abs/1604.08153 (2016).
3. Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6):26–38 (2017).
4. Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In Proc. of AAAI Conference on Artificial Intelligence, 1726–1734 (2017).
5. Helen C. Barron, Raymond J. Dolan, and Timothy E.J. Behrens. Online evaluation of novel choices by simultaneous representation of multiple memories. Nature Neuroscience, 16(10):1492–1498 (2013).
6. Lawrence W. Barsalou. Grounded cognition. Annual Review of Psychology, 59(1):617–645 (2008).
7. Daniel E Berlyne. Curiosity and exploration. Science, 153:25–33 (1966).
8. Benjamin Beyret, Ali Shafti, and A. Aldo Faisal. Dot-to-dot: Explainable hierarchical reinforcement learning for robotic manipulation. In Proc. of International Conference on Intelligent Robots and Systems (IROS), 5014–5019 (2019).
Dania Humaidan, Sebastian Otte, and Martin V. Butz. Fostering event composition using gated surprise. In Proc. of International Conference on Artiﬁcial Neural Networks (ICANN), 155–167 (2020).

David Isle, Mohammad Rostami, and Eric Eaton. Using task features for zero-shot knowledge transfer in lifelong learning. In Proc. of International Joint Conference on Artiﬁcial Intelligence (IJCAI), 1620–1626 (2016).

Max Jaderberg, Volodymyr Mnih, Wojciech Marian Czarnecki, Tom Schaul, Joel Z Leibo, David Silver, and Koray Kavukcuoglu. Reinforcement learning with unsupervised auxiliary tasks. International Conference on Learning Representations (ICLR). http://arxiv.org/abs/1611.05397 (2017).

M. Jeannerod. Mental imagery in the motor cortex. Neuropsychologia, 33 (11):1419–1432 (1995).

Yiding Jiang, Shixiang Gu, Kevin Murphy, Chelsea Finn, and Google Research. Language as an abstraction for hierarchical deep reinforcement learning. In Proc. of Neural Information Processing Systems (NeurIPS), 9414–9426 (2019).

Daniel Kahneman and Amos Tversky. The simulation heuristics. In Judgement under uncertainty: Heuristics and biases, 201–208 (1982).

Frédéric Kaplan and Pierre-Yves Oudeyer. Maximizing learning progress: an internal reward system for development. In Embodied Artiﬁcial Intelligence, 259–270 (2004).

Karl Josef Klaus. Teaching for analogical transfer as a means of improving problem-solving, thinking and learning. Instructional Science, 18(3):179–192 (1989).

David C. Knill and Alexandre Pouget. The bayesian brain: The role of uncertainty in neural coding and computation. Trends in Neuroscience, 22(2):712–719 (2004).

Tejas D. Kulkarni, Karthik R. Narasimhan, Arvadan Saeedi, and Joshua B. Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In Proc. of Neural Information Processing Systems (NIPS), 3675–3683 (2016).

Wilfried Kunde, Katrin Elsner, and Andrea Kiesel. No anticipation-no action: the role of anticipation in action and perception. Cogn Process, 8(2):71–78 (2007).

Gina R. Kuperberg. Tea with milk? A hierarchical generative framework of sequential event comprehension. Topics in Cognitive Science. doi: 10.1111/tops.12518 (2020).

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, 40:1–72 (2017).

George Lakoff and Mark Johnson. Philosophy in the flesh. Basic Books, (1999).

Guillaume Lample and Devendra Singh Chaplot. Playing fps games with deep reinforcement learning. AAAI Conference on Artiﬁcial Intelligence, 2140–2146 (2017).

Timothée Lesort, Natalia Díaz-Rodríguez, Jean-François Goudou, and Alexander C Li, Carlos Florensa, Ignasi Clavera, and Pieter Abbeel. Sub-task generalization with multi-task deep reinforcement learning. In Proc. of International Conference on Machine Learning (ICML), 3303–3313 (2018).
