13 Intrinsic Motivations for Open-Ended Learning

Gianluca Baldassarre

13.1 Introduction

Cognitive robotics and machine learning are producing a growing amount of works on *intrinsic motivations* (IMs) and *open-ended learning*. IMs, often contrasted to *extrinsic motivations* (EMs) that in animals are directed to satisfy biological needs such as hunger and thirst, refer to processes such as curiosity, surprise, novelty, and success at accomplishing one's own goals (Barto et al. 2004; Oudeyer et al. 2007; Baldassarre 2011; Baldassarre and Mirolli 2013). Open-ended learning refers to robots and agents that, similarly to the early development of humans (Weng et al. 2001; Lungarella et al. 2003), undergo prolonged periods of learning wherein they autonomously acquire knowledge and skills that might be useful to later solve tasks given by the user (Seepanomwan et al. 2017; Doncieux et al. 2018).

IMs are very important for robotics and machine learning because they can drive the autonomous open-ended learning of robots and machines by requiring little or no human intervention to furnish guidance in terms of data sets, behaviors to imitate, tasks, reward functions, and goals. Moreover, they allow the construction of robots and machines able to robustly operate in cluttered and unstructured environments posing challenges that cannot be anticipated at design time and preventing the possibility of programming behaviors in advance. Consider, for example, *service robots* that have to operate in warehouses, offices, houses, and health-care environments and in the fields of construction, agri-food, and space. Despite this importance, IMs are a subtle concept, as they come in different types, involve both functions ("what they are for") and mechanisms ("how do they work"), and can be mixed in various ways in the components of cognitive systems and robot controllers. This tends to generate quite a lot of confusion and to make it difficult to choose between the different available solutions when implementing robots and machines. This chapter addresses this problem in two ways. First (section 13.2), it provides computationally driven conceptual grids to define IMs by contrasting them with EMs and then to classify different types of IMs based on their possible functions and mechanisms, in particular by referring to three main classes of IMs here referred to as *epistemic intrinsic motivations* (eIMs). Second (section 13.3), it presents a selection of example models from cognitive robotics and machine learning to show how different IMs can be used to face different computational problems. The work concludes (section 13.4) by presenting some of the open challenges of the research on IMs.

13.2 Conceptual Grids: Mechanisms and Functions of Extrinsic and Intrinsic Motivations and Classes of (Epistemic) Intrinsic Motivations

The concept of IM has been proposed and developed within the psychological literature to overcome the difficulties of the behaviorist theory on learning and drives (e.g., Skinner 1938; Hull 1943), in particular to explain why animals spontaneously engage in puzzles (Harlow 1950) or can be instrumentally conditioned to produce particular responses on the basis of apparently neutral stimuli (e.g., a sudden light onset; Kish 1955), as happens with "standard" primary rewards (e.g., food). Subsequent proposals highlighted how the properties of certain stimuli can trigger animals' exploration and guide their learning processes—for example, when the stimuli are complex, unexpected, or in general surprising (Berlyne 1966). Another important thread of psychological research highlighted the importance that action plays in IMs—for example, in relation to the motivation coming from the fact that an agent manages to affect the environment with its behavior (*effectance*; White 1959) or can autonomously set its own goals and master their achievement (Ryan and Deci 2000). IMs involving actions are also related to sensorimotor contingencies studied by psychology and involving the mechanisms underlying the keen interest of animals and humans for the effects of their own actions (Polizzi di Sorrentino et al. 2014; Taffoni et al. 2014; Jacquey et al. 2019).

Within the computational sciences, Schmidhuber (1991a, 1991b) was the first to present a computational operationalization of some IM mechanisms (in particular *prediction-based IMs*; see below), and Barto et al. (2004) settled the fundamental link between IMs (in particular *competence-based IMs*; see below) and reinforcement-learning (RL) methods (Sutton and Barto 2018). These initial ideas were first developed within the developmental robotics scientific community (with works in the *IEEE Transactions on Autonomous Mental Development* journal, the International Conference on Development and Learning, and the Epigenetic Robotics Conference; Zlatev and Balkenius 2001; Lungarella et al. 2003; Oudeyer et al. 2007; Schembri et al. 2007; Doya and Taniguchi 2019), and more recently have been developed within the autonomous/cognitive robotics and machine-learning community (e.g., Bellemare 2016; Nair et al. 2018), in particular driven by the success of deep neural networks and RL (Goodfellow et al. 2017; Sutton and Barto 2018).

We now focus on understanding and defining these concepts more in detail and furnish conceptual grids on them. These grids are grounded in two perspectives from which one can look at cognitive processes (Tinbergen 1963; Marr and Poggio 1976): 1) the *computational functions* they serve—that is, the problems they solve: these indicate the possible "uses" for which they might be employed within an overall cognitive/robotic system; 2) the *mechanisms*, or algorithms: these refer to the information operations used to accomplish the functions. Some specifications are due on how the terms "functions" and "mechanisms" are used here. First, for animals "function" refers to *adaptive function*—that is, the utility of certain elements of intelligence, such as an IM, for the animal's biological fitness. For robots, "function" refers to the utility of a certain element of the robot's intelligence for the robot's *user*. Second, as with the functions in a computer program, "functions" can be organized at multiple hierarchical levels: from the highest level just mentioned ("biological fitness"; "utility for the user") to lower levels. For example, "moving an object as desired" can be further decomposed into lower-level functions such as "recognizing the object position" and "issuing suitable motor commands." Thus, a function can be seen as realized through a mechanism, but this mechanism in turn can be seen as a function to be realized with lower-level mechanisms. This downward decomposition can continue until some mechanisms are reached that are (arbitrarily) considered *primitive* for a given analysis.

13.2.1 Extrinsic and Intrinsic Motivations

What are motivations? Motivations are an element of intelligence having at least three important functions (for organisms; cf. Panksepp 1998): 1) *selection* drives the system to select a behavior, among alternative available ones, to attend the most important current needs/goals; 2) *energy* establishes the amount of energy invested in executing the selected behavior; 3) *learning* generates learning signals to change behavior. This chapter considers in particular the first and third functions of motivations. For example, we will see how IMs can drive an agent to move to some areas of the environment in navigation tasks (behavior selection) or can produce the reward signals for RL processes (production of learning signals).

What are *intrinsic motivations*? When initially studied in psychology, IMs were defined as motivations driving the performance of behavior "for its own sake"—that is, without any direct apparent purpose (Berlyne 1966). Although useful to guide intuition, this definition clarifies neither the functions nor the mechanisms of IMs. A more operational definition proposed here is that *intrinsic motivations are processes that can drive the acquisition of knowledge and skills in the absence of extrinsic motivations* (cf. Baldassarre 2011). IMs are hence best understood by contrasting them to *extrinsic motivations* (EMs). Table 13.1 highlights the main differences between EMs and a very important subset of IMs we will call *epistemic intrinsic motivations* (eIMs). In Baldassarre (2011) eIMs were considered to be IMs tout court, but here we recognize that they do not cover the full spectrum of

Table 13.1
Main features of extrinsic and (epistemic) intrinsic motivations (eIMs)

	Extrinsic motivations (EMs)	(Epistemic) intrinsic motivations (eIMs)
Function	Organisms: acquisition of *material resources*. Robots: accomplishment of *user's goals*.	Acquisition of *knowledge and skills*.
Mechanism	Organisms: measure the acquisition of *material resources* by getting information on their levels/changes *from body* and *resource monitoring*. Robots: measure the level/change of accomplishment of the *user's goals*.	Measure the acquisition of knowledge and skills by getting *information* on their levels/changes in *other parts of the brain (organisms)* or *controller (robots)*.
Time of contribution to the "ultimate" (extrinsic) function	*Immediately:* when the material resource is acquired and used (organisms); when the user's goals are accomplished (robots).	*Later:* when the acquired knowledge and skills are used to acquire resources (organisms) or to accomplish the user's goals (robots).
"Time signature" of the motivation	They tend to *go away* when the related resources are acquired and to *come back* when there is a lack of those resources.	They tend to *go away for good* when the related pieces of knowledge/skills are acquired.

IMs because, as we shall see, there are some IMs, which we call *other IMs* (oIMs), that are not eIMs. In the table, EMs are contrasted to eIMs because these form the core of IMs and because for their distinctive features they can help to clarify the overall nature of all IMs. The table entries illustrate this in more detail.

Regarding functions, EMs have the overall function of driving behavior and learning to the acquisition of material resources (Baldassarre 2011). For example, the EM of “hunger” drives behavior to look for and ingest food, and when this happens the behavior leading to it is strengthened. Instead, IMs have the overall function of driving behavior and learning toward the acquisition of knowledge and skills (note that “knowledge” also encompasses skills, but here “skills” are referred to explicitly to emphasize the aspects of knowledge more directly linked to action). For example, an IM related to novelty seeking could drive an agent to explore a novel object to learn its appearance, weight, shape, and so on. This function is shared by all IMs, not only by eIMs, as all IMs support the acquisition of knowledge and skills: in other words, all IMs have an epistemic function. In this respect, the term “epistemic motivations” might have been used in place of the term “intrinsic motivations,” which is somehow a misnomer as “intrinsic” suggests “internal” or at best, stretching it, “not directed to external material resources.” However, the term “intrinsic motivations” is kept here for its tradition. Moreover, the term IMs is handy to refer also to oIMs that, contrary to eIMs, are not based on an epistemic mechanism. In this respect, eIMs are the most prototypical IMs as they encompass *both* an epistemic function and an epistemic mechanism, and thus having a term that refers only to them is useful.

In terms of mechanisms, in animals EMs are based on measures of the acquisition of material resources by getting information on their levels/changes in the body or in the environment. For example, hunger, a drive guiding the selection of behaviors related to food seeking, might be triggered when the blood glucose level is low, and a reward-learning signal might be produced when food is ingested. Alternatively, an EM might be related to detecting the presence/availability of resources externally to the body—for example, the presence of a mating companion or the smell of prey in the environment (Baldassarre 2011). In robots, EMs are based on the measure of the accomplishment of the user’s goals; for example, a robot might self-charge its battery to remain operational and bring some objects to the user. Here the terms “extrinsic tasks/goals” will thus be referring to tasks/goals involving the acquisition of material resources or the accomplishment of the user’s goals. Incidentally, notice how EMs are a direct derivation of an evolutionary process not only for animals but also for robots: in animals, the acquisition of material resources is a means to increase biological fitness (number of fertile offspring) and, more specifically, the means for it—that is, survival and reproduction. Similarly, in robots the successful accomplishment of the user’s goals produces a higher chance that the specific features of the robot controller and physical structure are “reproduced,” as they are or in variants, in future robots.

Differently from EMs, eIMs rely on mechanisms that measure knowledge and skills by getting information on their levels/changes in other parts of the brain (for organisms) or in the controller (for robots). Importantly, this implies that an eIM involves the presence of at least three structures and functions inside the brain/controller (figure 13.1): (*a*) a *source component* that acquires knowledge; (*b*) an “*IM mechanism*” that measures the level or change of the knowledge of the source component; (*c*) a “*target component*” that receives the output of the IM mechanism and uses it to select behavior/energize behavior/drive learning processes. The core of this whole process is (*b*), the IM mechanism that measures the level or change of knowledge of the source component.

The specification above is very important, as, conceptually, eIMs involve the learning processes and knowledge of two different cognitive/computational components that might be very different in terms of the mechanisms and functions they play within the overall system, and this might make it difficult to recognize them in organisms or to recognize/implement them in robots. In some cases (figure 13.1*a*), the source component and the target component are the same data structure, in the sense that the IM mechanism detects the knowledge level/change in a component with the function of affecting the learning of the same component (possibly with the mediation of other components; figure 13.1*b*). For example, the selection of the skill to be trained among many skills to be learned might be based on the competence improvement of the skill itself (e.g., a robot might focus on learning to move one object, rather than on grasping it, if learning the first skill proceeds faster than for the second skill). In other cases (figure 13.1*c*), the source component and target components are distinct. For example, a component of a robot might detect the novelty of some objects, and this might drive a motor component to explore them with the function of improving its motor ability to manipulate them.

IMs that are not eIMs differ from the latter, as they do not use a learning source component as the origin of the motivation but rather other mechanisms: as anticipated, these

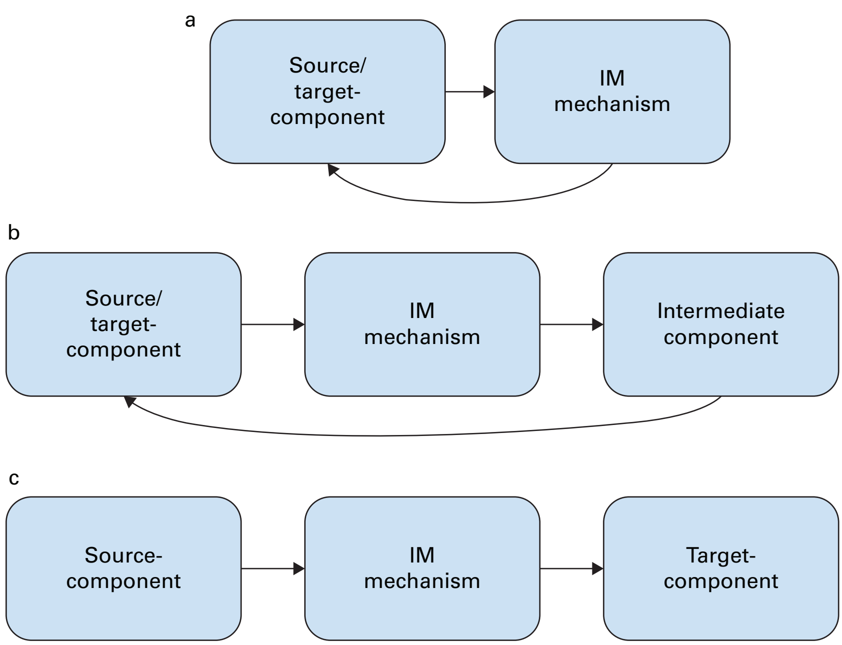

Figure 13.1
The key components of eIMs. (a) Case in which the source component and target component are the same structure. (b) Case in which the source component and the target component are the same structure, but the retroaction is mediated by an intermediate component. (c) Case in which the source component and the target component are different structures.

will be called *other IMs* (oIMs) to distinguish them from eIMs. Sometimes such "other mechanisms" mimic the acquisition of knowledge by a possible source component, but the latter is not actually present. For example, count-based novelty mechanisms (Bellemare et al. 2016) perform novelty detection on the basis of the frequency with which states are encountered rather than on the basis of how well they are memorized (although it is true that they are still present/absent in the counter memory). In other cases, other mechanisms are used that can support the function of acquiring knowledge and skills, but they themselves do not rely on a mechanism measuring the knowledge of some component. For example, the principle of empowerment (Klyubin et al. 2005), further discussed below, or the concept of bottlenecks (McGovern and Barto 2001), can support the acquisition of skills not by measuring the knowledge of a source component but by considering some properties of the environment or of the agent's actions.

A critical difference between EMs and all IMs is the time when they express their function—that is, their utility. EMs tend to express their function at a time very close to when they are triggered. This is because they lead to the acquisition and consumption of material resources (organisms) or to the accomplishment of the user's goals (robots), and when this happens they manifest their utility. Instead, IMs lead to the acquisition of knowledge and skills that are useful only later with respect to the time when they operate: the utility is indeed expressed only when such knowledge and skills are used to accomplish material resources or solve the user's goals.

The time when IMs and EMs express their utility is particularly important because it makes it difficult to actually measure the effectiveness of a given IM mechanism. A possible way to measure such effectiveness is to divide the life of the agent into two phases (Schembri et al. 2007; Baldassarre et al. 2019): 1) the *intrinsic motivation phase*, in which the agent uses IMs to acquire knowledge and skills without a direct utility; 2) the *extrinsic motivation phase*, in which the agent uses the knowledge and skills acquired in the intrinsic phase to solve extrinsic problems. These two phases resemble the two main phases of human life involving a first infancy/childhood phase, mainly guided by IMs, and an adulthood phase, mainly guided by EMs (Schembri et al. 2007). This idea of the two phases was set at the core of the REAL competition (Robot open-Ended Autonomous Learning; Cartoni et al. 2020) proposed to create a benchmark for open-ended learning. In this competition, during a first intrinsic phase a simulated camera-arm-gripper robot can freely interact with some objects to autonomously acquire knowledge and skills without being given any goal or reward; in a second extrinsic phase, the quality of such knowledge and skills is measured by asking the robot to solve some sampled extrinsic tasks involving the re-creation of some sampled object configurations. The robot's performance in the second phase thus furnishes a measure of the quality of the IM mechanisms used to acquire the knowledge in the first intrinsic phase. Two caveats come with this issue. Often in organisms, but also robots, IMs and EMs operate at the same time; for example, a robot might aim to learn how to manipulate an object while accomplishing a user's tasks. This requires suitable arbitration mechanisms to mediate between the time and resources dedicated to IMs and EMs. Second, IM and EM mechanisms and functions might be mixed. For example, a "source component" and an "IM mechanism" might support a "target component" pursuing an extrinsic goal. For example, the next sections show a common use of novelty-based IMs to improve exploration in the accomplishment of extrinsic RL tasks.

EMs and eIMs (and sometimes also oIMs) also have a typical “temporal signature” (Baldassarre 2011). In particular, EMs tend to go away when the resources they are directed at are obtained and to come back when such resources are consumed/lost. For example, hunger and the reward of food ingestion go away after a sufficient amount of food is ingested and, say, blood glucose level increases and come back when the blood glucose level is low again. Instead, eIMs triggered by the acquisition of a particular piece of information stored in the source component tend to go away forever when such a piece of information is acquired (unless the information is forgotten). From a cognitive perspective, this helps in recognizing whether a motivation is an (e)IM or an EM; from a computational perspective, this is relevant because it possibly generates nonstationary, challenging problems (e.g., a typical problem faced is that if an IM mechanism is used to produce a reward signal for an RL component, then the resulting reward function keeps changing and so should the behavior).

13.2.2 Three Classes of eIMs

The computational literature has greatly contributed to distinguishing between different classes of IM mechanisms. These classes in particular involve eIMs and often are not applicable to oIMs: the classification presented here uses the term “IMs” to stay with the common nomenclature, but it actually refers to eIMs. A first contribution (cf. Oudeyer and Kaplan 2007) distinguishes between knowledge-based IMs, related to the acquisition of information on the world, and competence-based IMs (CB-IMs), related to the acquisition of the capacity to act effectively. Another contribution (Barto et al. 2013) highlights the need to differentiate between two types of knowledge-based IMs—namely, novelty-based IMs (NB-IMs) and prediction-based IMs (PB-IMs), often confused within the computational and biological/cognitive literature. The main features of these three classes of IMs, summarized in table 13.2, are now considered in detail. The classes are based on the function

Table 13.2
The three classes of (e)IMs

	Novelty-based IMs	Prediction-based IMs	Competence-based IMs
Source component: nature	Memory component (pattern magazine)	Predictor (forward model)	Skill (inverse model)
Source component: function	Pattern storing and recoding	Prediction of patterns based on other patterns	Action selection
IM mechanism: type of knowledge measured	How well represented is the item in memory, or how much did its representation improve?	What is the prediction error or the prediction error change?	How efficient/effective is the skill to accomplish the task/goal?
IM mechanism: processes involved in the measurement	One process: memory check	Two processes: (a) prediction (b) comparison of prediction with data	Multiple processes: iterated perception-action performance, check of success
Target component: typical functions	- Store/recode new items - Direct attention - Drive physical exploration - Support goal formation	- Improve predictions - Drive physical exploration - Direct attention - Support goal formation	- Speed up the learning of multiple skills

Source: Partially based on Barto et al. 2013.

implemented by the source component. For each class, there exist many subclasses depending on the functions and mechanisms of the target component. The IM mechanism always measures the level or change of the knowledge of the source component.

NB-IMs are based on a memory source component that encodes patterns, such as percepts, with the function of storing and possibly recoding them in more useful formats—for example, to compress information or to facilitate downstream processes. The IM mechanism of NB-IMs measures knowledge of the source component based on a one-step process that checks the level of novelty/familiarity of a target pattern, such as an image from the world. Another possibility is that the IM mechanism measures the novelty change of the internal representation of the pattern, rather than its level: this can happen if the pattern is experienced multiple times and the source component progressively improves its representation. Typical functions realized by the target component involve storing/recoding novel items (which is the case when the source and target components coincide), directing attention to novel items, driving their physical exploration, or supporting goal formation.

PB-IMs are based on a predictor source component that predicts patterns on the basis of other patterns. In particular, the predictor receives as input a pattern, and possibly the agent's action, and on this basis predicts a target pattern in a future time. The "future time" involves a time range in which the target item should happen, but predictions can also be "in space," as in this example: "Given that I see a tree, I predict to see an apple if I look down 1 m." The IM mechanism of PB-IMs performs a measurement of the knowledge of the source component (predictor) based on a two-step process in which first the predictor predicts the target pattern on the basis of an input pattern, and possibly of the agent's action, and then the mechanism compares the prediction with the actual target pattern to compute the size of their mismatch—that is, to compute the prediction error. Another possibility is that the measure involves the prediction error improvement (change), rather than the prediction error (level), based on monitoring how the prediction error evolves in time. Typical functions played by the target component, possibly coincident with the source component, involve improving predictions, directing attention to unpredicted items, driving their physical exploration, and forming goals.

CB-IMs assume the existence of tasks/goals and are based on a skill source component that can accomplish the tasks/goals (e.g., within a given period of time, the trial). The skill is a closed-loop or open-loop controller (e.g., a dynamic movement primitive, a policy, or an option) potentially able to solve the task/achieve the goal. The IM mechanism of CB-IMs performs a measurement of the knowledge of the source component that involves a multistep process: 1) the skill acts to accomplish the task/goal, possibly based on multiple sensorimotor steps; 2) its competence level is measured, for example, in terms of the amount of reward collected during the trial, or in terms of goal achievement, or in terms of distance between the achieved state and the goal. Another possibility is that the IM mechanism actually measures the competence improvement, rather than the competence level, based on the monitoring of the performance at multiple times. CB-IMs are particularly important in cases where multiple skills for accomplishing different tasks/goals have to be learned. In this respect, typical functions realized by the target component, usually coincident with the source component, are to learn multiple skills/goals, and the IM mechanism speeds up their learning by focusing on the skills with the highest learning speed.

Note that the definition of CB-IMs assumes the existence of tasks/goals. This is a critical aspect of CB-IMs because open-ended learning agents should be able to autonomously generate or discover such tasks/goals, as these are a major means to learn skills in an incremental fashion (Mirolli and Baldassarre 2013). Various oIMs considered in the following sections can be used to support such self-generation/discovery of tasks/goals.

13.3 Cognitive Robotics and Machine-Learning Models

This section considers the main functions that can be supported by IMs through the presentation of some computational models drawn from the robotics and machine-learning literature. In particular, it focuses on how IMs serve the acquisition of the overall capacity of agents to interact in the world to modify it (Mirolli and Baldassarre 2013). This focus leads us to consider in particular the relation between IMs and RL, the learning paradigm most closely related to the acquisition of the capacity to act in the world. Given this focus, the IM functions considered here are as follows (figure 13.2): (*a*) the accomplishment of sparse extrinsic rewards; (*b*) the self-generation of goals; (*c*) the acquisition of skills, either as policies per se or as policies linked to goals. These functions in particular are accomplished through processes that rely strongly on IM mechanisms alongside other mechanisms; these other mechanisms are 1) exploration, 2) goal sampling, imagination, or "marking," and 3) the autonomous selection of skills to learn. Evolutionary processes are also considered to be possible general mechanisms searching the IM mechanisms themselves or the goals supporting CB-IMs.

13.3.1 Sparse Rewards

A first main function of IMs is to support the solution of RL tasks involving *sparse* extrinsic rewards—that is, rewards that are encountered rarely if the agent explores the environment randomly. Sparse rewards challenge learning agents, as they can be experienced only

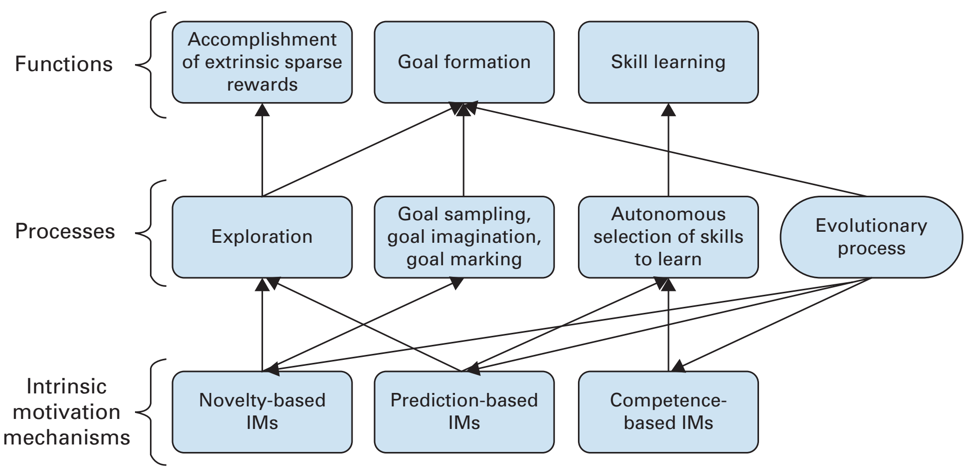

Figure 13.2
Some important functions that can be accomplished through IM mechanisms via some relevant processes.

after the performance of a long sequence of actions and therefore provide only very weak guidance for training. For example, imagine a camera-arm robot with no initial motor skills getting rewarded only for succeeding to grasp and lift an object with random movements. In this case it is close to impossible for a random exploration to lead to getting the reward and support learning. IMs can be very useful to solve tasks involving sparse rewards because they can facilitate the exploration of the environment through which the agent searches for the reward. Standard exploration methods, such as *ε-greedy* exploration (the agent selects a random action with a probability ε and the best action otherwise) and the Boltzmann distribution exploration (the possible actions are selected on the basis of a softmax function of their expected reward returns), are not adequate to face sparse-reward tasks because they lead to obtaining the reward only rarely. Various approaches have been proposed to produce a more effective exploration of the environment. A popular approach to foster exploration is based on NB-IMs. The idea is that the agent is attracted to states that it visited few times and tends to move away from familiar states. An extra reward (*novelty bonus*) could be given to the agent for making novel states attractive (Brafman and Tennenholtz 2002; Kakade and Dayan 2002). A nice property of novelty bonuses, and in general of IMs used to foster the pursuit of extrinsic rewards, is that since IMs have a transient nature, they tend to not affect the final policy acquired to maximize the final extrinsic reward.

A relevant class of methods using novelty to foster exploration in the search for extrinsic rewards is based on state novelty, measured as the number of times a state is encountered (Bellemare et al. 2016). In particular, these methods use density models to compute a pseudocount of the times in which states are visited based on the generalization of the counts for similar states. The method was successfully applied to agents able to solve the Atari game *Montezuma's Revenge*, involving a highly sparse reward. Another model used for similar purposes is presented in Burda et al. (2018). Here a random network is used to recode the state observations (images), and a second "copy" network is trained with supervised learning to "mimic" the first network (same input; desired output as the random network). The idea is that when states become more familiar the error of the copy network decreases.

Exploration to pursue extrinsic goals could also be pursued through PB-IMs. PB-IMs can rely on the prediction error (Schmidhuber 1991b) or the prediction error improvement (Schmidhuber 1991a) of a predictor network—that is, a world model predicting the next state on the basis of the current state and possibly the planned action. The prediction error has the disadvantage, if used by an IM mechanism, of not fading away in stochastic worlds. This problem is solved by the prediction error improvement, although at the cost of having a noisy and slow-adjusting signal. In the initial models using these strategies (Schmidhuber 1991a, 1991b), the predictor was used both as the source component and as the target component, meaning that the function of the used IM was to train the predictor itself. The same IM mechanism can, however, be also used to foster exploration to accomplish extrinsic tasks involving sparse rewards. An example of a model doing this is presented in Pathak et al. (2017). Here a forward model is used to produce a prediction error used as an intrinsic reward to train a RL agent to solve video games, such as *Mario Bros*., involving sparse extrinsic rewards. Interestingly, the model also proposes a mechanism to only focus on effects that are caused by the agent's actions by using a predictor that uses as input the internal representations of an inverse model predicting actions based on an input formed by the before-action state and the after-action state.

13.3.2 Goal Formation

A very interesting function for which IMs can be used is related to the acquisition of multiple sensorimotor skills that might be later used to accomplish other intrinsic tasks, or extrinsic tasks, particularly within a hierarchical RL framework where behavior is chunked into options (Sutton et al. 1999). Here we consider the goal-based version of options, in which each option involves (Barto et al. 2004; Singh et al. 2004) 1) a termination condition associated with the accomplishment of a goal, 2) an action policy indicating the primitive actions to select in correspondence with different states of the world, 3) possibly an initiation set encompassing the states from which if executed the policy is able to accomplish the goal. A goal is a representation of a set of world states that if reactivated internally drives the agent to act in the world so that the world assumes one of those states. There are various types of goals, such as goals as states of the world, goals as trajectories of states, avoidance goals, maintenance goals, and more (Merrick et al. 2016), but here we focus only on state goals for simplicity, and as many considerations can be extended to the other types of goals. Goals can have different levels of abstraction and can involve one's own body (Mannella et al. 2018; Hoffmann et al. 2010), the external environment (e.g., Santucci et al. 2016), the relation between a couple of elements (Kulkarni et al. 2016), or social aspects (Acevedo-Valle et al. 2018).

Various subfunctions, supported by IMs, are important for learning repertoires of multiple skills for later use. Here four are considered: 1) the autonomous generation of goals; 2) the coverage of the widest possible part of the goal space (goal exploration); 3) the generation of the reward for learning the policy of the single option; 4) the support of the progressive learning of skills, from easy to difficult, to speed up their acquisition.

The function of goal formation is important because during intrinsic learning the robots are not given any task to solve and so should autonomously self-generate tasks/goals guiding the acquisition of the related skills. Note that although goal formation is extremely important for open-ended learning, and various methods supporting it involve eIMs (Mirolli and Baldassarre 2013), it often also involves mechanisms differing from the elements of eIMs (source component, IM mechanism, and target component). These are here considered oIM mechanisms; further investigations are needed to understand if and how oIMs are linked to eIMs. We will now consider some relevant methods used to autonomously generate goals.

Goal sampling

When the goal space is given—for example, it is formed by the posture angles of a robot or the x, y positions of an object on a table—goals can be sampled on the basis of their skill learnability. For example, *goal babbling* (Rolf et al. 2010) allows a robot to self-generate posture goals that facilitate the learning of a coherent inverse model by maximizing the end-effector displacement, which favors the exploration of novel goals while minimizing the posture change, which favors the learning of regular versus awkward postures among the possible redundant postures. The approach has been later extended, for example, to learn multiple models in parallel (from end-effector position space to joint space and from the joint space to the motor space) through associative radial-basis-function networks growing on the basis of novel experiences (Rayyes and Steil 2019).

The goal space might not be given to the agent but form a subspace of the state or observation space to be actively searched. In this case goal sampling is not possible, especially if the subspace is small with respect to the whole space; in this case the goal subspace has to be actively discovered by the agent. Consider, for example, an observation space formed by images. In this case, the agent has to actively discover the image goals that it might actually achieve with its actions within the whole huge space formed by all possible images corresponding to all combinations of the pixel values. Now some approaches usable to this purpose are considered.

Goal marking

A number of models have proposed specific mechanisms to "mark"—that is, establish as goals—experienced states or observations. These models do not have the features of eIMs but can support open-ended learning via the formation of goals and the learning of the related skills, so they can be considered oIMs. A classic approach is the one for marking as goals the experienced states of the world that represent bottlenecks (McGovern and Barto 2001), nodal conditions that are often traversed when solving multiple extrinsic tasks (e.g., doorways when navigating an office).

Another model proposed to form goals corresponding to salient events, such as a change of light or sound (Barto et al. 2004; Singh et al. 2004). Linked to this, another approach proposed to mark as goals the novel observations that follow changes caused by the agent's actions in the environment (Santucci et al. 2016; Mannella et al. 2018). The idea behind this approach is that what robots (and organisms) ultimately should do during intrinsic learning is become able to change the world at will, so the observations that follow a change caused by own actions indicate a potential for doing this. The novelty of the changes guarantees that the goal has not been already formed. If changes in the world can also happen independently of the agent's action, additional mechanisms are needed to allow the agent to identify the subset of changes that depend on its action (Sperati and Baldassarre 2018; Pathak et al. 2017). Another approach forms goals when a particular relation between couples of elements takes place—for example, the "agent" picks up a "key" in an Atari game (Kulkarni et al. 2016).

A different approach (Zhao et al. 2012) uses RL to acquire various behaviors with motorized cameras within an active vision context (Ballard 1991; Ognibene and Baldassarre 2015)—for example, to lead two cameras to focus on the same target (vergence control). Here the model uses as a reward the accuracy of the reconstruction of images of a sparse-coding component (Olshausen and Field 1996), and the low error marks states where the two cameras manage to focus on the same target.

Another approach for skill learning is empowerment (Klyubin et al. 2005). Empowerment has a wide relevance for open-ended learning, but for lack of space only a few elements of it can be considered here. Empowerment is based on information theory and can be used to assign to each given world state a value that represents the variety of different outcome states that the agent can achieve with its actions from the given state. States with high empowerment can be used as target states; for example, their empowerment value can be directly used as reward to drive skill learning (T. Jung et al. 2011). Der and Martius (2015) propose another approach exploiting emergent properties of the environment-body-controller dynamics to autonomously acquire interesting motor skills in dynamic simulated agents. The skills are acquired on the basis of a simple two-layer neural network sensorimotor controller whose connection weights are trained through a differential extrinsic plasticity (DEP) rule derived from differential Hebbian learning (Zappacosta et al. 2018) that captures correlations between the changes of the input neurons and the output neurons.

Goal manifold search
This strategy searches goals within large observation spaces based on the idea that similar goals involve similar skills/actions, and so the performance of noisy variants of the already discovered skills/actions might possibly lead to discovering new achievable goals. This strategy was first used in a model (*skill babbling*; Reinhart 2017) to control an arm robot learning to displace an object in the 3D space. The model forms clusters of similar goals and discovers new goals by performing noisy versions of the actions corresponding to the centroid goals of clusters. The active goal manifold exploration model (AGME; Cartoni and Baldassarre 2018) actively discovers the goal manifold hidden in the observation space—for example, a posture space or an image space. For this purpose, the model builds a distance-based graph of the discovered goals, selects goals that have a higher distance from other discovered goals, generates perturbed versions of the policies associated with such goals, and performs them to discover new goals. The quality diversity algorithm (Kim et al. 2019) learns a repertoire of behaviors and goals by searching for behaviors that are different (novel) with respect to the already learned behaviors. The algorithm is, for example, used to allow a humanoid robot to acquire the skills to throw a ball into a basket located in many possible different positions (goals) on the floor. The hindsight experience replay approach (HER; Adrychowicz et al. 2017) exploits the outcome of policies to discover new goals, even if they are different from the pursued goal. The approach is very effective to incrementally discover new goals—for example, to manipulate objects in a simulated camera-arm-gripper robot.

Goal formation by imagination
Another related strategy discovers goals by first imagining them. For example, the reinforcement learning with imagined goals model (RIG; Nair et al. 2018), tested with a robot arm moving objects on a table, uses a generative model (a variational autoencoder; Kingma and Welling 2013) to first learn an internal compact representation of goals by randomly exploring the environment and then to "imagine" other possible goals whose skills are learned by RL. A later version of the model generates goals that have a high probability of being novel with respect to already learned goals by sampling them on the fringe of the distribution of the internal representation of the discovered goals (Pong et al. 2019). "Imagination" is a relevant means not only to generate goals but also to formulate plans to achieve those goals by assembling other goals/skills (Seepanomwan et al. 2015; Hung et al. 2018; M. Jung et al. 2019; Tanneberg et al. 2019) possibly acquired with IMs. This is an interesting trend that reformulates some high-level concepts elaborated by the classic symbolic planning literature (Russell and Norvig 2016), such as goals and planning, through neural network representations.

13.3.3 Selection of Skills to Train

The literature on animal learning (Skinner 1953) and on staged child development (Piaget 1953) shows that learning progress is faster if it proceeds from easy to difficult tasks. This strategy can also be used in artificial systems by training them with a curriculum involving increasingly difficult tasks (Asada et al. 1996; Bengio et al. 2009). One of the most interesting uses of IMs allows open-ended learning agents to autonomously select the skills needed to train to achieve goals possibly generated autonomously with the approaches illustrated above. Initially, PB-IMs were used to support the autonomous selection of tasks to learn (e.g., Singh et al. 2004; Oudeyer et al. 2007). Here the source component was a predictor, while the target component was the skill to learn, and the agent focused learning on skills causing the highest predictor error, or prediction error improvement, of the predicted skill outcome. Successively, CB-IMs were shown to be more appropriate than PB-IMs for selecting the skills to train because the predictor of the PB-IMs might learn to predict the skill outcome too early or too late with respect to when the controller finishes learning the skill. Instead, CB-IMs directly measure the competence acquired by different skills so it returns accurate information usable for selecting them (Santucci et al. [2013] compared these different IM mechanisms for task selection).

When a goal can be accomplished starting from a different initial condition, the CB-IM signal related to the goal must also take into account such an initial condition; moreover, when a goal can be selected not only depending on its learning rate but also depending on whether its achievement can be the precondition for learning other skills, then the CB-IM signal has to be used as a reward within a whole RL process selecting goals rather than actions (Santucci et al. 2019). IMs can also guide the progressive learning of increasingly difficult tasks represented at multiple levels of abstraction—for example, in robots learning to interact with different objects (Ugur and Piater 2016). In all these models, the skill of the selected goal should be trained (with RL) through a pseudo-reward equal to one when the goal is accomplished and to zero otherwise. This is more effective than what was done in the early years of research on IMs when the PB-IM signal used to select the goal/skill was also used to train the skill, as the PB-IM signal gradually fades away when the skill is learned.

13.3.4 Evolution

Tasks/goals could also be generated autonomously through evolutionary processes (genetic algorithms). Schembri et al. (2007) proposed the first model to do so in a population of RL simulated robots moving on a colored arena. During the intrinsic phase, the robots used intrinsic reward functions generated by a genetic algorithm to learn skills. In the later extrinsic phase, the robots learned to compose the acquired skills to accomplish extrinsic tasks (specific places in the arena). The success in learning these extrinsic tasks produced the fitness for the genetic algorithm. Singh et al. (2010) used an algorithm equivalent to evolution to search reward functions of RL agents engaged in searching for food in a grid world. They found that reward functions having the highest score rewarded the agents not only for searching for food but also for "opening boxes" where food was hidden. The model was used to suggest the existence of a continuum between EMs and IMs, rather than a distinction between them, as from an evolutionary perspective the two differ only for their distance from the events increasing fitness. The view proposed here distinguishes eIMs and EMs, as eIMs are based on the measure of *knowledge in a component of the controller,* whereas EMs are based on the measure of *material resources in the body or the environment*. It is, however, true that in the case of evolved oIMs that support the formation of goals and skills, as in the models reviewed above, a continuum with EMs can be seen since the criterion of the "knowledge-measurement" typical of eIMs is missing.

There is an additional important problem for open-ended learning that could be tackled with evolutionary approaches: Which goals/skills should be acquired, among those possible, to later best learn several different extrinsic tasks in a given domain? Del Verme et al. (2020) faced this problem and used a genetic algorithm to search goals/skills that were optimal for the solution of tasks drawn from a certain distribution of possible tasks in a given environment. The work showed how the optimal goals and skills depended on the time budget that the agent had in order to solve the extrinsic tasks and on the physical regularities of the environment. It so demonstrated that "fixed" mechanisms for goal generation, as those seen above, might lead to suboptimal solutions. Importantly, evolutionary approaches might thus be used to evolve the IM mechanisms themselves, as hinted by the arrows in figure 13.2 departing from the "evolutionary processes" box (Salgado et al. 2016). Although very interesting, this possibility is now limited by its high computational costs.

13.4 Conclusion

The study of intrinsic motivations is making important progress. However, many relevant open issues need further investigation. One open issue is the clarification of how non-epistemic intrinsic motivations work and are related to epistemic ones. Another open issue is the clarification of the link between intrinsic motivations and the autonomous formation of goals. A further issue, in part related to that, is the clarification of the relationship existing between intrinsic motivations and concepts such as empowerment and sensorimotor emergent behaviors. We have also seen how the computational literature is uncovering the existence of an articulated typology of intrinsic motivation mechanisms and functions. Understanding if and how these are also present in organisms' brains and behavior is a very interesting open problem.

Robot open-ended learning itself is still unsolved, as shown by the fact that we do not have robots able to undergo a truly open-ended learning experience leading to an unbounded accumulation of knowledge and skills. This might depend on multiple factors. On the side of goal formation, we have various mechanisms for the autonomous generation of goals, but all of them have limitations: goal sampling can only be applied to known small goal spaces; goal formation based on mechanisms such as bottlenecks, novel environment changes, goal-manifold discovery, and goal imagination has yet to be scaled to larger goal spaces and different domains. The autonomous selection of skills to train, based on competence-based intrinsic motivations, is becoming a standard, but it generally assumes discrete goals and hence must be further developed to be easily applicable to continuous goal spaces. Finally, systems working with discrete goals solve extrinsic problems based on planning and search methods that require the number of learned goal/skills to be limited to be efficient. This problem might be solved with evolutionary methods that indirectly search for a few robust skills to learn by searching the IM mechanisms themselves that lead to their generation; this, however, currently has a prohibitive computational cost.

Despite these challenges, the research field of open-ended learning driven by intrinsic motivations is surely one of the most exciting fields of cognitive robotics due to its potential for applications in robots acting in unstructured environments and to its close link with some of the most sophisticated and intriguing processes of human cognition, such as curiosity and the drive for the autonomous acquisition of knowledge.

Acknowledgments

This research received funding from the European Union under the 7th Research Framework Program, Grant Agreement n° 231722, Project "IM-CLeVeR—Intrinsically Motivated Cumulative Learning Versatile Robots"; and the Horizon 2020 Research and Innovation Program, Grant Agreement n° 713010, Project "GOAL-Robots—Goal-based Open-ended Autonomous Learning Robots."

Additional Reading and Resources

- A collection of works on intrinsic motivations and open-ended learning: Baldassarre, Gianluca, and Marco Mirolli. 2013. *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin: Springer.
- A work that complements the current work, with a perspective on the biology and brain mechanisms underlying intrinsic motivations: Baldassarre, Gianluca. 2011. "What Are Intrinsic Motivations? A Biological Perspective." In *Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011)*. New York: IEEE.
- A work presenting a general architecture supporting several of the functions for open-ended learning discussed in the chapter and usable for domains involving discrete goals: Santucci, Vieri Giuliano, Gianluca Baldassarre, and Marco Mirolli. 2016. "GRAIL: A Goal-Discovering Robotic Architecture for Intrinsically-Motivated Learning." *IEEE Transactions on Cognitive and Developmental Systems* 8 (3): 214–231. doi:10.1109/tcds.2016.2538961.
- Link to the project that sponsored this research, which furnishes additional resources and software on the topic: www.goal-robots.eu.

References

Acevedo-Valle, Juan M., Verena V. Hafner, and Cecilio Angulo. 2018. "Social Reinforcement in Artificial Prelinguistic Development: A Study Using Intrinsically Motivated Exploration Architectures." *IEEE Transactions on Cognitive and Developmental Systems* 12 (2): 198–208.

Andrychowicz, Marcin, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, Pieter Abbeel, and Wojciech Zaremba. 2017. "Hindsight Experience Replay." ArXiv preprint: 1707.01495v3.

Asada, Minoru, Shoichi Noda, Sukoya Tawaratsumida, and Koh Hosoda. 1996. "Purposive Behavior Acquisition for a Real Robot by Vision-Based Reinforcement Learning." *Machine Learning* 23 (2–3): 279–303.

Baldassarre, Gianluca. 2011. "What Are Intrinsic Motivations? A Biological Perspective." In Vol. 2, *2011 IEEE International Conference on Development and Learning*, 1–8. New York: IEEE.

Baldassarre, Gianluca, William Lord, Giovanni Granato, and Vieri Giuliano Santucci. 2019. "An Embodied Agent Learning Affordances with Intrinsic Motivations and Solving Extrinsic Tasks with Attention and One-Step Planning." *Frontiers in Neurorobotics* 13 (45): e1–e26.

Baldassarre, Gianluca, and Marco Mirolli. 2013. *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin: Springer.

Ballard, Dana H. 1991. "Animate Vision." *Artificial Intelligence* 48:57–86.

Barto, Andrew, Marco Mirolli, and Gianluca Baldassarre. 2013. "Novelty or Surprise?" *Frontiers in Psychology* 4 (907): e1–e15.

Barto, Andrew G., Satinder Singh, and Nuttapong Chentanez. 2004. "Intrinsically Motivated Learning of Hierarchical Collections of Skills." In *Proceedings of the 3rd International Conference on Developmental Learning (ICDL2004)*, 112–119. New York: IEEE.

Bellemare, Marc G., Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. 2016. "Unifying Count-Based Exploration and Intrinsic Motivation." ArXiv preprint: 1606.01868.

Bengio, Yoshua, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. "Curriculum Learning." In *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. New York: Association for Computing Machinery.

Berlyne, Daniel E. 1966. "Curiosity and Exploration." *Science* 153 (3731): 25–33.

Brafman, Ronen I., and Moshe Tennenholtz. 2002. "R-Max—a General Polynomial Time Algorithm for Near-Optimal Reinforcement Learning." *Journal of Machine Learning Research* 3:213–231.

Burda, Yuri, Harrison Edwards, Amos Storkey, and Oleg Klimov. 2018. "Exploration by Random Network Distillation." ArXiv preprint: 1810.12894v1.

Cartoni, Emilio, and Gianluca Baldassarre. 2018. "Autonomous Discovery of the Goal Space to Learn a Parameterized Skill." ArXiv preprint: 1805.07547v1.

Cartoni, Emilio, Francesco Mannella, Vieri Giuliano Santucci, Jochen Triesch, Elmar Rueckert, and Gianluca Baldassarre. 2020. "REAL-2019: Robot Open-Ended Autonomous Learning Competition." In *Proceedings of Machine Learning Research* 1:142–152.

Del Verme, Manuel, Bruno Castro Da Silva, and Gianluca Baldassarre. 2020. "Optimal Options for Multi-task Reinforcement Learning under Time Constraints." ArXiv preprint: 2001.01620v1.

Der, Ralf, and Georg Martius. 2015. "Novel Plasticity Rule Can Explain the Development of Sensorimotor Intelligence." *Proceedings of the National Academy of Science USA* 112 (45): e6224–e6232.

Doncieux, Stephane, David Filliat, Natalia Díaz-Rodríguez, Timothy Hospedales, Richard Duro, Alexandre Coninx, Diederik M. Roijers, Benoît Girard, Nicolas Perrin, and Olivier Sigaud. 2018. "Open-Ended Learning: A Conceptual Framework Based on Representational Redescription." *Frontiers in Neurorobotics* 12 (59): e1–e6.

Doya, Kenji, and Tadahiro Taniguchi. 2019. "Toward Evolutionary and Developmental Intelligence." *Current Opinion in Behavioral Sciences* 29:91–96.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2017. *Deep Learning*. Cambridge, MA: MIT Press.

Harlow, Harry F. 1950. "Learning and Satiation of Response in Intrinsically Motivated Complex Puzzle Performance by Monkeys." *Journal of Comparative and Physiological Psychology* 43 (4): 289–294.

Hoffmann, Matej, Hugo Marques, Alejandro Arieta, Hidenobu Sumioka, Max Lungarella, and Rolf Pfeifer. 2010. "Body Schema in Robotics: A Review." *IEEE Transactions on Autonomous Mental Development* 2 (4): 304–324.

Hull, Clark L. 1943. *Principles of Behavior*. New York: Appleton-Century-Crofts.

Hung, Chia-Chun, Timothy Lillicrap, Josh Abramson, Yan Wu, Mehdi Mirza, Federico Carnevale, Arun Ahuja, and Greg Wayne. 2018. "Optimizing Agent Behavior over Long Time Scales by Transporting Value." ArXiv preprint: 1810.06721v1.

Jacquey, Lisa, Gianluca Baldassarre, Vieri Giuliano Santucci, and J. Kevin O'Regan. 2019. "Sensorimotor Contingencies as a Key Drive of Development: From Babies to Robots." *Frontiers in Neurorobotics* 13 (98): e1–e20.

Jung, Minju, Takazumi Matsumoto, and Jun Tani. 2019. "Goal-Directed Behavior under Variational Predictive Coding: Dynamic Organization of Visual Attention and Working Memory." ArXiv preprint: 1903.04932v1.

Jung, Tani, Daniel Polani, and Peter Stone. 2011. "Empowerment for Continuous Agent Environment Systems." *Adaptive Behavior* 19 (1): 16–39.

Kakade, Sham, and Peter Dayan. 2002. "Dopamine: Generalization and Bonuses." *Neural Networks* 15 (4–6): 549–559.

Kim, Seungsu, Alexandre Coninx, and Stephane Doncieux. 2019. "From Exploration to Control: Learning Object Manipulation Skills through Novelty Search and Local Adaptation." ArXiv preprint: 1901.00811v1.

Kingma, Diederik P., and Max Welling. 2013. "Auto-Encoding Variational Bayes." ArXiv preprint: 1312.6114v10.

Kish, George Bela. 1955. "Learning When the Onset of Illumination Is Used as the Reinforcing Stimulus." *Journal of Comparative and Physiological Psychology* 48 (4): 261–264.

Klyubin, Alexander S., Daniel Polani, and Chrystopher L. Nehaniv. 2005. "All Else Being Equal Be Empowered." In *European Conference on Artificial Life*, 744–753. Berlin: Springer.

Kulkarni, Tejas D., Karthik Narasimhan, Ardavan Saeedi, and Josh Tenenbaum. 2016. "Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation." In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 3682–3690. New York: Currant Associates.

Lungarella, Max, Giorgio Metta, Rolf Pfeifer, and Giulio Sandini. 2003. "Developmental Robotics: A Survey." *Connection Science* 15 (4): 151–190.

Mannella, Francesco, Vieri Giuliano Santucci, Somogyi Eszter, Lisa Jacquey, Kevin J. O'Regan, and Gianluca Baldassarre. 2018. "Know Your Body through Intrinsic Goals." *Frontiers in Neurorobotics* 12:30.

Marr, David, and Tomaso Poggio. 1976. “From Understanding Computation to Understanding Neural Circuitry.” *A.I. Memo AIM–357*. Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Marraffa, Rodolfo, Valerio Sperati, Daniele Caligiore, Jochen Triesch, and Gianluca Baldassarre. 2012. “A Bio-inspired Attention Model of Anticipation in Gaze-Contingency Experiments with Infants.” In *2012 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics*, 1–8. New York: IEEE.

McGovern, Amy, and Andrew G. Barto. 2001. “Automatic Discovery of Subgoals in Reinforcement Learning Using Diverse Density.” Faculty Publication Series. University of Massachusetts Amherst, Computer Science Department.

Merrick, Kathryn, Nazmul Siddique, and Inaki Rano. 2016. “Experience-Based Generation of Maintenance and Achievement Goals on a Mobile Robot.” *Journal of Behavioral Robotics* 7 (1): 67–84.

Mirolli, Marco, and Gianluca Baldassarre. 2013. “Functions and Mechanisms of Intrinsic Motivations: The Knowledge versus Competence Distinction.” In *Intrinsically Motivated Learning in Natural and Artificial Systems*, 49–72. Berlin: Springer.

Nair, Ashvin, Vitchyr Pong, Murtaza Dalal, Shikhar Bahl, Steven Lin, and Sergey Levine. 2018. “Visual Reinforcement Learning with Imagined Goals.” In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 9209–9220. New York: Currant Associates.

Ognibene, Dimitri, and Gianluca Baldassarre. 2015. “Ecological Active Vision: Four Bio-inspired Principles to Integrate Bottom-Up and Adaptive Top-Down Attention Tested with a Simple Camera-Arm Robot.” *IEEE Transactions on Autonomous Mental Development* 7 (1): 3–25.

Olshausen, Bruno, and David J. Field. 1996. “Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images.” *Nature* 381 (6583): 607–609.

Oudeyer, Pierre-Yves, and Frederic Kaplan. 2007. “What Is Intrinsic Motivation? A Typology of Computational Approaches.” *Frontiers in Neurorobotics* 1:6.

Oudeyer, Pierre-Yves, Frédéric Kaplan, and Verena V. Hafner. 2007. “Intrinsic Motivation Systems for Autonomous Mental Development.” *IEEE Transactions on Evolutionary Computation* 11 (2): 265–286.

Panksepp, Jaak. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford: Oxford University Press.

Pathak, Deepak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. 2017. “Curiosity-Driven Exploration by Self-Supervised Prediction.” In *The IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 488–489. New York: IEEE.

Piaget, Jean. 1953. *The Origins of Intelligence in Children*. London: Routledge and Kegan Paul.

Polizzi di Sorrentino, Eugenia, Gloria Sabbatini, Valentina Truppa, Anna Bordonali, Fabrizio Taffoni, Domenico Formica, Gianluca Baldassarre, Marco Mirolli, and Visalberghi Guglielmelli Eugenio. 2014. “Exploration and Learning in Capuchin Monkeys (Sapajus Spp.): The Role of Actionoutcome Contingencies.” *Animal Cognition* 17 (5): 1081–1088.

Pong, Vitchyr H., Murtaza Dalal, Steven Lin, Ashvin Nair, Shikhar Bahl, and Sergey Levine. 2019. “Skew-Fit: State-Covering Self-Supervised Reinforcement Learning.” ArXiv preprint: 1903.03698v2.

Rayyes, Rania, and Jochen Steil. 2019. “Online Associative Multi-stage Goal Babbling toward Versatile Learning of Sensorimotor Skills.” In *2019 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics*, 327–334. New York: IEEE.

Reinhart, Felix. 2017. “Autonomous Exploration of Motor Skills by Skill Babbling.” *Autonomous Robots* 41 (7): 1521–1537.

Rolf, Matthias, Jochen J. Steil, and Michael Gienger. 2010. “Goal Babbling Permits Direct Learning of Inverse Kinematics.” *IEEE Transactions on Autonomous Mental Development* 2 (3): 216–229.

Russell, Stuart J., and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Harlow, UK: Pearson Education.

Ryan, Richard M., and Edward L. Deci. 2000. “Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.” *American Psychologist* 55 (1): 68–78.

Salgado, Rodrigo, Abraham Prieto, Pilar Caamaño, Francisco Bellas, and Richard J. Duro. 2016. “Motiven: Motivational Engine with Sub-Goal Identification for Autonomous Robots.” In *IEEE Congress on Evolutionary Computation*, 4887–4894. New York: IEEE.

Santucci, Vieri Giuliano, Gianluca Baldassarre, and Emilio Cartoni. 2019. “Autonomous Reinforcement Learning of Multiple Interrelated Tasks.” In *2019 Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, 221–227. New York: IEEE.

Santucci, Vieri Giuliano, Gianluca Baldassarre, and Marco Mirolli. 2013. “Which Is the Best Intrinsic Motivation Signal for Learning Multiple Skills?” *Frontiers in Neurorobotics* 7 (22): e1–e14.

Santucci, Vieri Giuliano, Gianluca Baldassarre, and Marco Mirolli. 2016. "GRAIL: A Goal-Discovering Robotic Architecture for Intrinsically-Motivated Learning." *IEEE Transactions on Cognitive and Developmental Systems* 8 (3): 214–231.

Schembri, Massimiliano, Marco Mirolli, and Gianluca Baldassarre. 2007. "Evolving Childhood's Length and Learning Parameters in an Intrinsically Motivated Reinforcement Learning Robot." In *the 7th International Conference on Epigenetic Robotics (Epirob2007)*, 141–148. Lund, Sweden: Lund University Press.

Schmidhuber, Juergen. 1991a. "Curious Model-Building Control Systems." In *International Joint Conference on Artificial Neural Networks*, 1458–1463. Piscataway, NJ: IEEE.

Schmidhuber, Juergen. 1991b. "A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers." In *International Conference on Simulation of Adaptive Behavior: From Animals To Animats*, 222–227 Boston: MIT Press.

Schmidhuber, Juergen. 2010. "Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010)." *IEEE Transactions on Autonomous Mental Development* 2 (3): 230–247.

Seepanomwan, Kristsana, Daniele Caligiore, Angelo Cangelosi, and Gianluca Baldassarre. 2015. "Generalization, Decision Making, and Embodiment Effects in Mental Rotation: A Neurorobotic Architecture Tested with a Humanoid Robot." *Neural Networks* 72:31–47.

Seepanomwan, Kristsana, Vieri Giuliano Santucci, and Gianluca Baldassarre. 2017. "Intrinsically Motivated Discovered Outcomes Boost User's Goals Achievement in a Humanoid Robot." In *2017 Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, 178–183. New York: IEEE.

Singh, Satinder, Andrew G. Barto, and Nuttapong Chentanez. 2004. "Intrinsically Motivated Reinforcement Learning." In *Advances in Neural Information Processing Systems*, 1281–1288. Boston: MIT Press.

Singh, Satinder, Richard L. Lewis, Andrew G. Barto, and Jonathan Sorg. 2010. "Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective." *IEEE Transactions on Autonomous Mental Development* 2 (2): 70–82.

Skinner, Burrhus F. 1938. *The Behavior of Organisms*. New York: Appleton-Century-Crofts.

Skinner, Burrhus F. 1953. *Science and Human Behavior*. New York: Macmillan.

Sperati, Valerio, and Gianluca Baldassarre. 2018. "A Bio-inspired Model Learning Visual Goals and Attention Skills through Contingencies and Intrinsic Motivations." *IEEE Transactions on Cognitive and Developmental Systems* 10 (2): 326–344.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Sutton, Richard S., Doina Precup, and Satinder Singh. 1999. "Between MDPS and Semi-MDPS: A Framework for Temporal Abstraction in Reinforcement Learning." *Artificial Intelligence* 112:181–211.

Taffoni, Fabrizio, Eleonora Tamilia, Valentina Focaroli, Domenico Formica, Luca Ricci, Giovanni Di Pino, Gianluca Baldassarre, Marco Mirolli, Eugenio Guglielmelli, and Flavio Keller. 2014. "Development of Goal-Directed Action Selection Guided by Intrinsic Motivations: An Experiment with Children." *Experimental Brain Research* 232 (7): 2167–2177.

Tanneberg, Daniel, Jan Peters, and Elmar Rueckert. 2019. "Intrinsic Motivation and Mental Replay Enable Efficient Online Adaptation in Stochastic Recurrent Networks." *Neural Networks* 109:67–80.

Tinbergen, Nikolaas. 1963. "On Aims and Methods of Ethology." *Zeitschrift fur Tierpsychologie* 20 (4): 410–433.

Ugur, Emre, and Justus Piater. 2016. "Emergent Structuring of Interdependent Affordance Learning Tasks Using Intrinsic Motivation and Empirical Feature Selection." *IEEE Transactions on Cognitive and Developmental Systems* 9 (4): 328–340.

Vieira Neto, Hugo, and Ulrich Nehmzow. 2007. "Visual Novelty Detection with Automatic Scale Selection." *Robotics and Autonomous Systems* 55 (9): 693–701.

Weng, Juyang, James McClelland, Alex Pentland, Olaf Sporns, Ida Stockman, Mriganka Sur, and Esther Thelen. 2001. "Autonomous Mental Development by Robots and Animals." *Science* 291:599–600.

White, Ruth W. 1959. "Motivation Reconsidered: The Concept of Competence." *Psychological Review* 66:297–333.

Zappacosta, Stefano, Francesco Mannella, Marco Mirolli, and Gianluca Baldassarre. 2018. "General Differential Hebbian Learning: Capturing Temporal Relations between Events in Neural Networks and the Brain." *PLoS Computational Biology* 14 (8): e1006227.

Zhao, Yu, Constantin A. Rothkopf, Jochen Triesch, and Bertram E. Shi. 2012. "A Unified Model of the Joint Development of Disparity Selectivity and Vergence Control." In *2012 IEEE Joint International Conference on Development and Learning and Epigenetic Robotics*, 1–6. New York: IEEE.

Zlatev, Jordan, and Christian Balkenius. 2001. "Introduction: Why Epigenetic Robotics?" In *Proceedings of the First International Workshop on Epigenetic Robotics*, 1–4. Lund, Sweden: Lund University Press.