Functions and mechanisms of intrinsic motivations

The knowledge vs. competence distinction

Marco Mirolli and Gianluca Baldassarre

Istituto di Scienze e Tecnologie della Cognizione,
Consiglio Nazionale delle Ricerche,
Rome, Italy
{marco.mirolli,gianluca.baldassarre}@istc.cnr.it

Abstract. Mammals, and humans in particular, are endowed with an exceptional capacity for cumulative learning. This capacity crucially depends on the presence of intrinsic motivations, i.e. motivations that are not directly related to an organism's survival and reproduction but rather to its ability to learn. Recently, there have been a number of attempts to model and reproduce intrinsic motivations in artificial systems. Different kinds of intrinsic motivations have been proposed both in psychology and in machine learning and robotics: some are based on the knowledge of the learning system, while others are based on its competence. In this contribution we discuss the distinction between knowledge-based and competence-based intrinsic motivations with respect to both the functional roles that motivations play in learning and the mechanisms by which those functions are implemented. In particular, after arguing that the principal function of intrinsic motivations consists in allowing the development of a repertoire of skills (rather than of knowledge), we suggest that at least two different sub-functions can be identified: (a) discovering which skills might be acquired and (b) deciding which skill to train when. We propose that in biological organisms knowledge-based intrinsic motivation mechanisms might implement the former function, whereas competence-based mechanisms might underly the latter one.

1 Introduction

The capacity of autonomous cumulative learning demonstrated by complex organisms like mammals, and humans in particular, is astonishing. This capacity is likely to have its roots in *intrinsic motivations*, i.e. motivations not directly related to extrinsic rewards such as food or sex, but rather to what the animal knows (curiosity, novelty, surprise) or can do (competence). Both animal and human psychologists have found evidence indicating that intrinsic motivations play an important role in animals' behavior and learning (Berlyne, 1960; Deci, 1975; Deci and Ryan, 1985; White, 1959).

Recently, the study of intrinsic motivations has been gaining increasing attention also in machine learning and robotics, as researchers in these fields have recognized that truly intelligent artificial systems need to develop their own abilities while autonomously interacting with their environment (Weng et al., 2001). The potentially open-ended complexification of a system's skills might require the use of learning signals that are non-task-specific, and hence intrinsic. As a result, several computational models of intrinsically motivated learning have been proposed so far, but the study of artificial intrinsically motivated cumulative learning systems is still in its infancy.

The aim of the present chapter is to aid the development of this field by clarifying issues related to different kinds of intrinsic motivations, both with respect to the possible mechanisms that might implement intrinsic motivations and to the possible functions that they might play in cumulative learning. In particular, the paper focuses on the distinction between knowledge-based and competence-based intrinsic motivations, that is intrinsic motivations that are based on what the system *knows* vs. on what the system *can do*. Note that what the system knows can include also knowledge (e.g. predictions) about the results of the systems actions. And, viceversa, competence-based intrinsic motivations might involve the use of predictions for obtaining measures related to competence. What really distinguish knowledge-based by competence-based system is that the former use measures that are related to the capacity of the system to model its environment (including the system's body and the interactions between the system and the environment), whereas the latter use measure that are related to the system's ability to have specific effects on the environment.

We start (section 2) by reviewing the psychological literature on intrinsic motivations, highlighting an important distinction that can be made between intrinsic motivations driven by what a system knows (knowledge-based) and those driven by what it can do (competence-based). Then we review the computational modeling literature from the perspective of an analogous distinction between knowledge-based and competence-based intrinsic motivation systems (section 3). In section 4 we suggest that the distinction between knowledge-based and competence-based systems can and should be made not only with respect to the *mechanisms* of intrinsic motivations, but also with respect to their *functions*, that is with respect to what kind of learning they support; and we argue that the ultimate function of intrinsic motivation is to support the cumulative learning of skills, rather than knowledge. Unfortunately, this is not the case for many of the computational models proposed in the literature so far (section 5), in particular for those employing knowledge-based mechanisms. Indeed, we suggest that purely knowledge-based systems might not be particularly well suited for driving the cumulative acquisition of skills (section 6). Finally, we consider the problem of intrinsic motivations from the point of view of a hierarchical learning system (section 7): here we argue that different kinds of intrinsic motivations might play different functional roles at different levels of the hierarchy, and we refer to the possible neural bases of both knowledge-based and competence-based intrinsic motivations in real brains. Section 8 concludes the paper by summarizing our contributions and discussing promising directions for future research.

2 Intrinsic motivations in Psychology

Interest in intrinsic motivations first arose in the 1950s in the field of animal psychology, as several researchers discovered a number of phenomena that were in apparent contrast to the extremely influential theory of motivation proposed by Hull (1943). According to Hull's theory, animal behavior was motivated by drives, conceived as temporal physiological deficits that the organism is led to reduce in order to achieve homeostatic equilibrium. Typical examples of drives are hunger and thirst, which make the animal work for achieving respectively food and water so as to satisfy its needs for energy and liquid. According to Hull, all motivations are either physiological primary drives or secondary drives derived from primary ones through learning (as it happens in classical conditioning experiments).

Notwithstanding the popularity of Hull's theory, soon animal psychologists reported phenomena that were difficult to reconcile with the drive concept. For example, Harlow and co-workers (Harlow, 1950; Harlow et al., 1950) reported that rhesus monkeys might spend long periods of time in trying to solve mechanical puzzles without any reward. Kish and colleagues reported that operant conditioning phenomena (i.e., phenomena related to the fact that the rate of responding to a manipulandum can be influenced by the consequences of these responses) could be elicited in mice not only by primary rewards but also by apparently neutral stimuli that had never been associated with rewards, such as "microswitch clicks, relay noises, and stimuli produced by a moving platform" (Kish, 1955; Kish and Antonitis, 1956). Similarly, Butler showed that rhesus monkeys could learn a discrimination task by using as a reinforcement just the possibility to look at other conspecifics from a window (Butler, 1953).

Some authors, such as Montgomery (1954), tried to reconcile these findings with Hull's theory by postulating the existence of other drives, like drives to manipulate, to play, or to explore. But this move was hard to accept because exploratory drives do not seem to possess any of the two fundamental characteristics of primary drives: they are not related to any internal deficit and they do not seem to have any homeostatic function.

2.1 Knowledge-based views

The considerations reported above led several psychologists to develop explanations of intrinsic motivations that were not based on the drive concept. Among them, probably the most influential proposal was the one of Berlyne (1960). According to him animal exploration and intrinsically motivated activities in general depend on the fact that animals are attracted by optimal levels of novelty of stimuli, of their complexity, and of surprise, conceived as a falsification of the animal's expectations. As it can be noted, all the intrinsic motivations proposed by Berlyne are knowledge-based, in that they are related to the properties of the stimuli that the animal perceives and on their relation to the animal's knowledge. (Note that Berlyne never made the distinction between knowledge and competence we are discussing here: this is just our interpretation of his work.)

Several similar knowledge-based proposals have been made in the psychological literature. These postulated either that animals are motivated to receive an optimal level of stimulation (or of novelty of stimuli) (e.g. Hebb, 1955; Hunt, 1965) or that they are motivated to reduce the discrepancy (or incongruity, or dissonance) between their knowledge and their current perception of the environment (e.g. Dember and Earl, 1957; Festinger, 1957; Kagan, 1972).

2.2 Competence-based views

On the other hand, in his seminal review on motivations, White (1959) strongly advocated a competence-based view of intrinsic motivations, proposing that animals have a fundamental motivation for *effectance*, i.e. for the capacity to have effective interactions with their environment. White's paper had a great influence on subsequent research on motivation. In particular, in the fields of educational and human psychology the link between intrinsic motivations and the concept of competence has remained quite strong since then (Deci, 1975). For example, according to *self-determination theory* (Deci and Ryan, 1985) there is a continuum between extrinsically and intrinsically motivated activities. Among the most important factors that make an activity intrinsically motivated are the sense of *autonomy*, that is the perception that the activity is self-determined, and the sense of *competence*, that is the perception that we have (or are getting) mastery of the activity (Ryan and Deci, 2000). Similarly, De Charms proposed that personal causation, that is the sense of having control over one's environment, was a fundamental driving force of human behavior (De Charms, 1968). Likewise, the *theory of flow* of Csikszentmihalyi postulates that humans are motivated to engage in activities that represent an appropriate level of learning challenge, i.e. that are neither too easy nor too difficult to master given the individual's current level of competence (Csikszentmihalyi, 1991).

3 Computational models of intrinsic motivations

The distinction between knowledge-based and competence-based intrinsic motivations can also be appropriately applied within the context of the computational literature on intrinsic motivations. In this field, a useful and typical way of framing the problem of intrinsic motivations consists in considering it within the computational framework of reinforcement learning (e.g Barto et al., 2004; Oudeyer and Kaplan, 2007; Schembri et al., 2007c; Schmidhuber, 1991b, see also Barto, 2012). Reinforcement learning algorithms are developed so as to maximize the sum of future rewards, where a reward is defined as a numerical value that is continuously received by the learning agent (Sutton and Barto, 1998). In this context, intrinsic motivations can be conceived as components of the rewards that are not directly related to the task that the agent must solve but are rather task-independent.

3.1 Knowledge-based models

Most of the proposed models of intrinsic motivations are knowledge-based as they depend on the *stimuli* perceived by the learning system (and on their relations with the system's expectations, including those related to the results of the system's actions) rather than on the system's *skills*. For example, the first model of intrinsic motivations for an artificial reinforcement learning agent was proposed by Schmidhuber (1991b). It consisted in adding to a reinforcement learning agent an adaptive world model that learned to predict the next perception given the current perception and the planned action, and in using the errors in these predictions as an intrinsic reward for the system. The intrinsic reward complemented the extrinsic reward related to the task at hand. The rationale of this proposal was that in this way the "curious" reinforcement learning agent would be pushed not only to maximize external rewards but also to improve its own world model, thus exploring poorly-known parts of the environment.

The recognition that this kind of system would get stuck in parts of the environment that are unpredictable (where the prediction error can never decrease) led Schmidhuber to propose to use as an intrinsic reward signal a measure of the learning progress (namely, the decrease in prediction error) of the world model (Schmidhuber, 1991a). An analogous idea has been recently explored in the context of developmental robotics by Oudeyer and colleagues under the name of Intelligent Adaptive Curiosity, together with a mechanism for automatically dividing the whole sensorimotor space into subregions within which to compute the learning progress and on which to focus learning (Oudeyer et al., 2007). Several other models have been presented in which the intrinsic motivation consists of some form of perceived novelty, prediction error, or learning progress of a world model (Huang and Weng, 2002; Lee et al., 2009; Marshall et al., 2004; Merrick and Maher, 2009; Saunders and Gero, 2002; Storck et al., 1995).

3.2 Competence-based models

Contrary to what has been stated by Oudeyer and Kaplan in their useful review (Oudeyer and Kaplan, 2007), there are also a few computational models of artificial systems whose learning is driven by some form of competence-based intrinsic motivations. The first of such systems is the Intrinsically Motivated Reinforcement Learning (IMRL) agent proposed by Barto and colleagues (Barto et al., 2004; Stout et al., 2005), which is based on the reinforcement learning framework of *options* (Sutton et al., 1999). In short, an option is a temporally extended action defined by an initiation set (the set of states in which the option can be initiated), a policy (the states-actions mapping followed during the execution of the option), and a termination set (the set of states in which the option terminates). In practice, an option defines a skill that can be recalled each time the agent needs to achieve one of the option termination states. These states can be considered as the option goals. In IMRL options are created for reaching different *salient events*, where salient events are specified by the experimenter as the opening of a door, the ringing of a bell, or the switching on of a light. Intrinsic rewards, perceived when the salient events are reached, are defined as $1-$ the (estimated) probability of reaching the salient event with the executed option. In other words, the intrinsic reinforcement is given on the basis of the unpredicted occurrence of salient events. As the authors themselves explicitly state, this is a form of prediction error and so, in this respect, it is analogous to the knowledge-based intrinsic motivations discussed above. However, the prediction error in this case is not calculated at every state, but only on the goal states of the options (whenever these are reached). Hence, the intrinsic reward is given on the basis of the (lack of) ability of the system to reach one of its goal-states. Hence, in this case the intrinsic reward should be considered as competence-based, rather than knowledge-based.

The use of competence-based intrinsic rewards is more explicit in a model we have proposed (Schembri et al., 2007a,b,c). This was directly inspired by the IMRL framework, but instead of using options it implements skills on the basis of *experts*, each of which is an instance of the actor-critic reinforcement learning architecture (Barto et al., 1983) trained with the temporal differences (TD) learning algorithm (Sutton, 1988). The model is hierarchical and so it is also formed by a *selector* which assigns the control of action (and the possibility of learning) to experts. Each expert is assigned a reward function and when given the control it improves the policy of its actor (states-actions mapping) on the basis of the *TD error*, that is the error of its critic related to the prediction of future discounted rewards. The selector learns to assign the control to experts on the basis of an intrinsic reward signal represented by the TD error of the expert selected at each step. The rationale of this is that the TD error of an expert can be considered as a good estimate of the improvement of its policy. Indeed, a positive average TD error of an expert means that it is behaving better than expected by its own critic, which means that it is improving its ability to maximize rewards. Viceversa, a zero average TD error indicates that it is behaving as expected, meaning that it is not improving. By receiving the TD error of the selected expert as its reward, the selector will learn to give the control to the expert that, in a given context, is expected to improve its skill the most.

Here it is clear, even more than in Barto's IMRL, that even though the intrinsic motivation signal for the selector (the TD error of the selected expert) is a form of prediction error, it is a competence-based signal, since the prediction error is relative to the obtained rewards, and hence it is a measure of skill improvement rather than of the agent's ability (or inability) in modeling its environment. The fact that the TD error is a signal that is related to performance has important consequences for the possible problem of unlearnability. As discussed above, an intrinsically motivated knowledge-based system that uses the prediction error as a reward will get stuck in parts of the environment that are not predictable. In contrast, a system that uses the TD error as its intrinsic reward will not get stuck in parts of the environment where the reward is unpredictable because in that case the evaluations (predictions of future discounted rewards) will become equal to the average received reward and the TD error will be sometimes positive (when the reward is received) and sometimes negative (when the reward is not received). Hence, the reinforcement learning system will learn to predict that the average reward is zero and will pass to explore other parts of the environment where some skill learning progress can be made. In fact, in contrast to the method proposed by Barto and colleagues, this kind of competence-based intrinsic motivation is able to solve not only the problem of unpredictability of rewards, but also the problem of unlearnability of a skill, and for the same reason. If, for whatever reason, a skill cannot be learned, the evaluations of the expert will equal the average received reward, the TD error signal will average to zero, and the selector will learn that the skill cannot be trained and will prefer experts that can improve their performance.

Two more recent models of competence based intrinsic motivations are Stout and Barto (2010), which is quite similar to our own, and Baranes and Oudeyer (2010), which is not based on reinforcement learning but on control theory.

4 Functions vs. mechanisms and the primacy of competence over knowledge

Up to now we have reviewed the psychological and computational modeling literature on intrinsic motivations from the point of view of the distinction between knowledge-based and competence-based hypotheses. Following Oudeyer and Kaplan (2007) we have so far related this distinction to the *mechanisms* that drive intrinsically motivated behavior and/or learning processes. An analogous but even more important distinction can be done with respect to the *function* that intrinsic motivations are considered to play within the overall system. Whereas there is a general agreement on the fact that intrinsic motivations serve the role of driving the learning of a system, it is much less clear (but it is important to clarify) which kind of learning intrinsic motivations are supposed to support: is it learning of *knowledge* or learning of *competence*? In other words: do intrinsic motivations help the system in building *increasingly accurate models* or do they allow to discover *more effective ways of acting on* (interacting with) the environment?

The answer will probably reflect a more general and fundamental (we might call it *philosophical*) attitude that one takes in considering human beings and other animals as well as in building artificial intelligent systems. What is more important: knowledge or action? No doubt, the western culture has been giving prominence to knowledge over action: from Plato's *World of Ideas*, to Descartes' *cogito ergo sum*, to Kant's primacy of the Critics of *Pure Reason* over the Critics of Practical Reason, and so on. This strong cultural legacy is evident also in modern *Cognitive* Science, which poses at the very center of the study of psychology the study of how human beings process information inside their heads rather than of how they behave and interact with their environment. Since their birth in the 1950s, the same general attitude has also informed the *sciences of the artificial* (Simon, 1996), that is disciplines such as artificial intelligence and machine learning. For decades the focus of research of these disciplines has been on reasoning, problem solving, knowledge representation on one hand (Russell and Norvig, 2003), and perception and categorization on the other (Mitchell, 1997), whereas the study of behavior has been almost neglected. This state of affairs has started to change, at least from the point of view of artificial systems, in the 1990s, when classical artificial intelligence was criticized and "invaded" by new robotics (Brooks, 1991), artificial life (Langton, 1989), adaptive behavior research (Meyer and Wilson, 1990), and reinforcement learning (Sutton and Barto, 1998), which started to gain increasing attention within machine learning. Indeed, what is common to all these new approaches and clearly distinguishes them from the old ones is their attitude with respect to the knowledge versus competence issue: all of them emphasize the primacy of behavior over representation, of action over cognition, and of competence over knowledge (Clark, 1997; Pfeifer and Scheier, 1999).

The rationale behind this fundamental shift lies in evolutionary considerations: animals evolved on the basis of their capacity to survive and reproduce, and survival and reproduction depend primarily on what organisms *do*, not on what they *know*. What is selected for is the capacity to adaptively interact with one's own environment in ever increasingly complex and efficient ways, not the capacity to accurately represent the environment in one's head (on this point, see also Barto, 2010). This does not mean that the capacity to represent the world is useless and should not be investigated. Rather, it means that knowledge is ancillary to behavior, rather than the other way round. Hence, in general, the capacity to model the environment should be considered as one possible (but not necessary) tool for improving an agent's capacity of interacting with the environment, and not as the ultimate goal of an organism or of an artificial system. In other words, *some* form of knowledge might be useful in *some* circumstances for *some* purposes. But it should be kept in mind that perfect knowledge is not the ultimate goal of any living organism, and it should not be the goal of any useful artificial system. Indeed, understanding which kind and level of knowledge can be useful for which kind of behaving system in which circumstances is a fundamental challenge for both the empirical and the synthetic behavioral and brain sciences.

5 Functional roles of intrinsic motivations in computational models

Coming back to the question about the ultimate function of intrinsic motivation which we introduced in the previous section, it should now be clear what our stance is, and why: whatever the *mechanisms* which implement intrinsic motivations, their primary and fundamental *function* is to allow a system to develop a cumulative repertoire of useful skills (rather than an increasingly accurate knowledge of the environment). We think not only that this is true for real organisms, but that this should be the standpoint from which to consider the endeavor of building intrinsically motivated artificial systems. We feel the need to underscore this point because it is not clear to what extent this position is shared among computational modelers.

Surprisingly as it might seem, the majority of computational models of intrinsic motivations not only use knowledge-based *mechanisms*, but have the acquisition of knowledge as their only *function*. For example, Schmidhuber's (1991a) curious agent is a "model builder", whose goal is the improvement of its predictions. The same is true for the robots used to test the Intelligent Adaptive Curiosity algorithm of Oudeyer and collegues (Oudeyer et al., 2007), and for almost all other systems whose intrinsic motivations are based on knowledge (e.g. Huang and Weng, 2002; Lee et al., 2009; Marshall et al., 2004; Merrick and Maher, 2009). On the contrary, competence-based intrinsic motivation mechanisms have been always used for improving skills (Baranes and Oudeyer, 2010; Barto et al., 2004; Schembri et al., 2007c; Stout and Barto, 2010).

It is important to note that this need not necessarily be the case. Indeed, the distinction between the functions of knowledge-based and competence-based intrinsic motivations is different from the one related to the mechanisms underlying them. It is possible for knowledge-based intrinsic motivation mechanisms to help improve the learning agent's skills (as it happens, for example, in Schmidhuber, 2002), as it is possible for competence-based mechanisms to serve the acquisition of knowledge. But these possibilities in principle must be demonstrated in practice, and in doing so there might be some difficulties. As the focus here is in the use of intrinsic motivations for the acquisition of skills, the next section discusses only the possible problems related to the use of knowledge-based mechanisms for accumulating competence, while it does not consider the less relevant possibility of using competence-based intrinsic motivations for acquiring knowledge.

6 Knowledge-based mechanisms and competence

As a representative example of a purely knowledge-based system (i.e. a system which uses only knowledge-based mechanisms) let us consider the work of Oudeyer et al. (2007). In the presented experiments, the authors show how the proposed system demonstrates an interesting developmental trajectory: it starts by focusing its learning on those parts of the sensorimotor space where predictions are easier, and when knowledge in those parts has been acquired the system shifts its attention to the parts where predictions are more difficult. But while it is clear that through this developmental history the system accumulates increasingly complex *knowledge* about the sensorimotor contingencies of its world (and of the robot's body - world interactions), the authors never show that any *competence* (i.e. ability to *do* something) has been acquired. All you can observe is that the actions that are chosen at the beginning of learning are different from those that are chosen at the end, but *no accumulation of skills is demonstrated.* The same happens in most other works on knowledge-based intrinsic motivation systems (e.g. Huang and Weng, 2002; Lee et al., 2009; Marshall et al., 2004; Schmidhuber, 1991b; Storck et al., 1995, but see Baranes and Oudeyer, 2009 for a case in which the use of the knowledge acquired through intrinsic motivations for control, i.e. competence, *is* demonstrated).

What are then the relationships between such knowledge-based systems and the acquisition of competence? There might be different answers to this question, the first of which simply stating that the ultimate goal of learning is not to acquire skills, but rather knowledge. For the reasons discussed above, however, we are not satisfied by this position. Indeed, if one is interested in understanding the mechanisms underlying the intelligent *behavior* of organisms, or in building robots which can *act* intelligently, one has to solve the problem of how knowledge can support the acquisition of skills.

One possibility is that once an accurate model of the sensory consequences of actions have been acquired, this model can directly guide behavior through its *inversion*, as it is done in some forms of model-based control (Hof et al., 2009): given the current goal of the system and the current context, the agent might internally simulate the consequences of all its actions and choose to perform the one whose predicted consequence are more similar to the goal (this is the approach taken, for example, by Baranes and Oudeyer, 2009). This approach has the problem that it may be computationally very expensive, and it might become just unfeasible in realistic conditions, where the spaces of states and actions are high-dimensional or continuous. In these conditions, as primitive actions cannot usually directly lead to desired states, sequences of actions must be used and so the complexity of the search grows exponentially.

Another possibility might be to suggest that the acquired model can be used for training the behavioral controller, as it happens, for example, in the *distal teacher* framework (Jordan and Rumelhart, 1992). Notwithstanding the attractiveness of this idea, this seems to represent a very indirect route to skill learning. Indeed, consider that this solution would involve (a) training of a behavioral system (on the basis of knowledge-based intrinsic motivations) so that (b) it favors the training of a predictor (c) which then is used for training a second behavioral system (or for retraining the first one) to pursue its goals. Why not following a more straightforward route and directly train the skills? We do not want to deny that building models of the effects of the agent's own actions is useful for learning kills. In fact, we do think that prediction plays fundamental roles for action execution and learning. What we are questioning here is the idea that models are learnt *before* and *independently from* the skills that they are supposed to support. Hence, in our view, modeling abilities should *complement* and *aid* acquisition of skills, not *replace* them.

Yet another possibility is to claim that using knowledge-based intrinsic rewards can directly help the system not only in improving its knowledge, but also in improving its skills (this is the position taken by Schmidhuber, 2012). The idea is that by being rewarded by some form of prediction error (or improvement in prediction) a behaving system will tend to reproduce the actions that lead to unexpected consequences. Once the capacity to systematically reproduce those consequences has been acquired, a good model can be developed, prediction error (improvements) will decrease together with the intrinsic rewards, and the system will move on and try to learn something else. Though we think that this idea is on the right track, it might still suffer of some important limitations.

The fundamental problem is that in order for a purely knowledge-based mechanism to optimally support the acquisition of a cumulative collection of skills, one should assume a very strict and direct link between knowledge and competence that seems difficult to justify. For example, what can assure that the predictions related to the system's own actions are in general more interesting than those that depends only on the environment? If this cannot be guaranteed, what would prevent a purely knowledge-based intrinsic-motivation system to stop moving and spend its time in passively observing and modeling interesting parts of the environment *instead of* focusing on learning useful skills?

Furthermore, there are certainly a lot of situations that are challenging to predict but for which the actions are relatively trivial or even irrelevant. For example, imagine a humanoid robot that has the goal of clearing a table full of different objects by pushing them on the floor. Predicting the effects of own actions is surely possible but very complex. Instead, the action of pushing all objects on the floor is rather easy to be learned and accomplished. If the system is intrinsically motivatedy only to increase its knowledge what would prevent the system in spending its time moving all the objects in all possible positions instead of focusing on the simple actions that would satisfy the goal? Of course, here the level of detail of the predicion is of the most importance, as for example predictions regarding the presence or absence of objects on the table (rather tha their exact location) would be easier to learn and suitable to drive the acquisition of the skill of clearing the table. But the problem is that which is the right level of detail for the prediction *depends on the competence you want to acquire*; how can a purely knowledge-based system decide which is the level of abstraction that is appropriate for improving competence?

And for many difficult skills there might even be no level of detail in predictions that would be appropriate for the the system to learn. If the problem is too difficult, the desired consequences might be just too unlikely to manifest for the skill to be learned through knowledge based intrinsic motivations alone. For example, think about the ability to juggle three balls: how could one acquire this ability through purely knowledge-based intrinsic motivations? If predictions are done at the level of the positions of the balls in space the system would go on forever in learning the irrelevant predictions about all possible results of its actions of throwing the balls in all directions. But even if the predictions regard a more appropriate representation of the state space, for example reagrding whether or not a ball has been caught, the task is so difficult that after some unsuccessfull attempts, the predictor will learn that the behaving system is not able to catch the ball on the fly, the prediction error (and prediction improvements) will go to zero, and the system will loose interest in the activity and pass to something else. The general point is that when I want to learn juggling my behavior seems to be driven by a motivation that is directly related to the acquisition of that particular skill, and not just to the general predictions of the consequences of my movements.

Finally, another key problem of the models of knowledge-based intrinsic motivations proposed so far is related to the cumulativity of learning. When the behaving system has learnt to systematically reproduce a certain consequence and the rewarding prediction improvements decrease to zero, the reinforcement learning system will begin to receive less rewards than expected and start to *unlearn* the just acquired behavior. This is exactly what happens, for example, in the system of Schmidhuber (1991b): once predictions within a given part of the sensory-motor space are learnt, the system gets bored and learns to *avoid* the actions just acquired. Hence, after training the system has cumulated knowledge but no competence. For an intrinsically motivated system to be cumulative in the acquisition of skills, it is necessary not only to temporarily focus its training on different sensorimotor activities, but also to have a means for storing skills after they have been successfully discovered so that the system's repertoire can increase and old skills can be reused and combined to learn new ones. Hence, it seems impossible to study intrinsic motivations for cumulative learning without considering the architectural issue of how skills can be cumulated.

7 Intrinsic motivations and skills accumulation

7.1 Hierarchy and modularity of skill organization

A multi-task learning system seems to need two key ingredients: (a) some form of structural modularity, where different skills are at least partially stored in different parts of the system, and (b) some form of hierarchical organization, where a higher-level system is responsible for selecting which low-level module (expert) to use and train in each context. The need for a structurally modular and hierarchical organization of action has sometimes been questioned, most notably by the work of Botvinick and Plaut (2004) and of Tani and colleagues (e.g. Tani et al., 2008; Yamashita and Tani, 2008). But the two basic features of structural modularity and hierarchy are common to the vast majority of computational models that deal with the acquisition of several skills, even though they are instantiated in different ways within different systems (Baldassarre, 2002; Barto et al., 2004; Barto and Mahadevan, 2003; Caligiore et al., 2010; Dayan and Hinton, 1993; Dietterich, 2000; Doya et al., 2002; Haruno et al., 2001; Parr and Russell, 1997; Schembri et al., 2007c; Singh, 1992; Tani and Nolfi, 1999; Wiering and Schmidhuber, 1997).

Furthermore, action selection and learning in real brains seem in fact to display a considerable level of modular and hierarchical organization (Fuster, 2001; Grafton and Hamilton, 2007; Miller and Cohen, 2001; Redgrave et al., 1999). Action selection is supposed to be implemented in the basal ganglia (Mink, 1996; Redgrave, 2007), a group of subcortical structures that is also thought to be responsible for the processes underlying reinforcement learning phenomena (Barto, 1995; Doya, 2000; Houk et al., 1995; Joel et al., 2002). Interestingly, the basal ganglia form a number of parallel loops with most of the cortex, and in particular with those parts that are most directly related to action, i.e. the whole frontal cortex and associative parietal and temporal areas (Alexander et al., 1986; Joel and Weiner, 1994). These parallel loops can be observed at different levels of abstractions. First, loops are present at the level of macro-areas, thus forming different networks supposed to implement different functions: e.g. a limbic network (with orbitofrontal cortex), an associative network (with dorsal prefrontal cortex, parietal cortex, and temporal cortex), and a sensory-motor network (with motor and premotor cortex) (Yin and Knowlton, 2006). Second, parallel loops are present *within* each macro-area, for example distinguishing parts of the motor cortex related to the control of different actuators, e.g. an arm (Romanelli et al., 2005). Finally, *within* each sub-area dedicated to a specific actuator, different loops might be responsible for implementing different actions (Gurney et al., 2001).

If one assumes that the accumulation of skills requires a learning architecture that is at least partially modular and hierarchical, the problem of identifying good intrinsic learning signals splits in two sub-problems, as different levels of the hierarchy are likely to require different signals. In a system composed of several experts implementing different skills and a selector that arbitrates between them, the problem for each expert consists in identifying which skill to acquire and how, whereas the problem for the selector consists in deciding what to learn and when, i.e. which skill to train in each context (Baldassarre and Mirolli, 2012). What kind of learning signals can help to solve these problems? We argue that knowledge-based rewards might be used for training the experts, while competence-based learning signals might be used for training the selector.

7.2 Knowledge-based signals for skill acquisition

In line with most machine learning research, here we consider a skill as the ability to reach a certain final state (goal) in some conditions. This is also in line with how behavior seems to be organized in real brains, where actions are defined, at least at the cortical level, by their end-states, i.e. their outcomes, rather than by the specific movements that are performed by the subject (Graziano, 2006; Rizzolatti and Luppino, 2001). In this respect, probably the most important problem that an autonomous learning system should solve is the identification of the skills that it is worth acquiring. In other words, an autonomous cumulative learning agent must not only learn how to reach certain states, but also which states of the world can be achieved in certain contexts given the appropriate behavior. We think that knowledge-based intrinsic motivation mechanisms might represent appropriate solutions to both these problems. In particular, both the discovery of which are the possible skills to be acquired and the training of these skills might be driven by the same reinforcement signals provided by the detection of *unpredicted* events.

A possible way in which this might happen has been recently proposed by Redgrave and Gurney (2006, see also Gurney et al., 2012; Redgrave et al., 2012) as the repetition bias hypothesis. If the detection of an unpredicted change in the environment is immediately followed by a learning signal that biases the agent to reproduce the actions performed before the change, this might lead the system to focus learning on the change and to identify which are the specific aspects of the behavior that caused the event. This result will be achieved on the basis of the statistics of the agent-environment interactions: those aspects of the behavior that cause the environmental change will be reinforced more than those that do not, thus further increasing the probability of being deployed. As the behavior of the system is sculpted and starts to reliably produce the same environmental change, the change becomes predictable on the basis of the performed actions, and so stops producing intrinsic rewards (Mirolli et al., tted; Santucci et al., 2010).

This hypothesis can explain otherwise puzzling findings related to the neural basis of reinforcement learning, in particular, regarding the phasic activation of the neuromodulator dopamine. Phasic dopamine is released in the basal ganglia when rewards like food are received, and this enhances the plasticity of striatal synapses (Reynolds and Wickens, 2002). Furthermore, after learning has taken place, phasic dopamine is no more released in conjunction with reward delivery, but is rather anticipated at the time of the presentation of the stimulus that allowed predicting the reward (conditioning). If after learning the reward is omitted, a dip in baseline dopamine release is observed at the time of the predicted reward. All these data have led to propose that phasic dopamine constitutes the biological analog of the TD error signal of reinforcement learning models, and represents the reward prediction error that drives biological reinforcement learning (Houk et al., 1995; Schultz et al., 1997; Schultz and Dickinson, 2000).

But there are several known phenomena that do not fit the reward prediction error hypothesis of phasic dopamine, which are not discussed here as they are reviewed at length in Redgrave et al. (2012). The most important point for the present discussion is the fact that phasic dopamine is triggered not only in the presence of rewards (or reward predicting stimuli, see above), but also in response to all kinds of unpredicted salient events, including sudden luminance changes and sounds (Dommett et al., 2005; Horvitz, 2000), likely based on the processes taking place in a subcortical brain structure named *superior colliculus*. This supports the hypothesis discussed above that unpredicted events can generate intrinsic reinforcement signals that support the acquisition of novel actions. Interestingly, dopamine release is triggered not only by the superior colliculus in response to unexpected simple environmental changes, but also by the *hippocampus* in response to the detection of novel situations (Lisman and Grace, 2005, see also Otmakova et al., 2012). Indeed, the novelty signals of the hippocampus seem to depend on the detection of associative novelty, that is novelty in the contextual, spatial, and temporal association between perceived stimuli. For these reasons we speculate that the release of dopamine in the presence of novel situations might underlie the intrinsically motivated learning of skills of increasing complexity, where the outcomes of the actions to be acquired are not only defined by simple instantaneous events but also by spatial and temporal relationships between objects in the environment.

It is important to note that these dopamine signals that are supposed to drive the intrinsically motivated acquisition of skills are simple signals related to unpredicted (or novel) events and states, *not to prediction learning progress*, as predicted by sophisticated knowledge-based intrinsic motivation models (Oudeyer et al., 2007; Schmidhuber, 1991a, see also (Schmidhuber, 2012)). How then can biological intrinsic motivation systems avoid getting stuck in trying to learn and reproduce events that are not predictable? The solution can be in the presence of competence-based intrinsic motivation mechanisms at the higher level of the hierarchy.

7.3 Competence-based signals for deciding which skill to train

If, as discussed above, cumulative learning depends on a hierarchical system where different modules learn different skills and a higher-level system selects which module must control behavior in each moment, then the unpredictability problem need not be addressed at the level of the single skills, but can be solved at the level of the selector. Indeed, as the selector is in charge of making strategic decisions about the system activities, it is the selector that has to avoid wasting time in situations where there is nothing to be learned. And this can be done in case the selector is trained with some form of competence-based intrinsic reward signal.

A working example is the use of the TD error of the experts as the selector reward in the modular hierarchical reinforcement learning system presented in section 3 (Schembri et al., 2007a,b,c). As mentioned there, such an intrinsically motivated system does not risk getting stuck in unlearnable situations as the TD learning of experts used to train the selector is a measure of the *learning progress* of the experts themselves, so the selector will learn to use, in each context, the expert that is learning the most. This works equally well whatever the reason why an expert cannot learn: an ability too difficult to be acquired or intrinsic stochasticity of action effects and rewards. In any case, the selector will use an expert only as long as it can learn something; when no competence learning progress is made, the TD error will average to zero and the selector will prefer to move to something else.

Importantly, in order for such a solution to work, the competence-based intrinsic reward to the selector must measure learning progress of a skill (as it is the case for Schembri et al., 2007a,c and for the more recent model of Stout and Barto, 2010). If the reward reflects the inability of an expert to reach its goal state, as in the original proposal of intrinsically motivated reinforcement learning of Barto et al. (2004), the overall system is subject to the problem of unlearnability just as a purely knowledge-based system trained by prediction errors. In this case, the system could easily get stuck trying to pursue states (or changes of states) that do not depend on its own actions, or that are just completely random.

While the biological knowledge-based signal that is supposed to drive the learning of the single skills can be identified with the release of phasic dopamine on the basis of unexpected events (and possibly with the information about associative novelty originating in the hippocampus), the identification of the possible biological implementation of a competence-progress intrinsic motivation signal is a completely open issue. As the signal should be related to the improvements in the acquisition of the agent's ability to systematically produce some environmental change, then it should originate in parts of the brain that may monitor the outcomes of the agent's actions (and, as we are talking about intrinsically motivated learning, the possible action outcomes should not be innately provided with value). Finally, since these signals should drive the reinforcement learning of the high-level selector, it might be related to dopamine, as the lower-level learning signal.

Interestingly, the pre-limbic cortex, a part of the prefrontal cortex of the rat, might have the prerequisites to be involved in such a competence-progress evaluation. First, it is a brain region that is known to be involved in goal-directed action-outcome learning (Dalley et al., 2004; Heidbreder and Groenewegen, 2003). Second, it receives a direct input from the hippocampus, where signals regarding the novelty of situations are generated. Third, it represents one of the major excitatory inputs of the ventral tegmental area dopaminergic neurons (Geisler et al., 2007), and the activation of these connections are known to produce dopaminergic release (Taber et al., 1995). As we are not aware of any hypothesis about the functional role of the connections from the prelimbic cortex to the dopaminergic neurons, it would be interesting to investigate the possibility that they underlie competence-based intrinsic motivation signals.

8 Summary and open challenges

In this contribution we have discussed the different functions that intrinsic motivations might play in cumulative learning, and the different mechanisms that may support these functions in both natural and artificial learning systems. With respect to the *function*, we have argued that the ultimate role of intrinsic motivations is supporting the cumulative acquisition of increasingly complex skills, and that computational research on this issue would significantly benefit from making this point more explicit in the proposed models: in our view, models of intrinsic motivations should aim at letting artificial agent increase their *skills*, with the eventual accumulation of *knowledge* playing only a supporting role.

With respect to the *mechanisms*, we have argued that both knowledge-based and competence-based mechanisms might play important roles. The cumulative acquisition of skills might require some sort of modular and hierarchical architecture, in which different experts acquire different skills and a selector decides which expert to use in each situation. In such a hierarchical learning system, intrinsic motivations might play two sub-functions: (1) driving the learning of the experts by setting which are the skills that are to be acquired and (2) driving the decisions of the selector regarding which expert to train in each context. On the basis of biological evidence, we have suggested that knowledge-based mechanisms like the detection of unpredicted events might support the former function, while we have speculated that competence-based mechanisms measuring the progress in skill acquisition might support the latter. Several important challenges need to be tackled in future research, in particular with respect to the goal of modeling the intrinsically motivated cumulative learning of real organisms.

A first interesting research direction has to do with knowledge-based intrinsic motivations. In particular, with the intrinsic reinforcements that are supposed to support the acquisition of single skills through the discovery of which are the environmental changes that depend on the behavior of the learning agent and that thus can become target outcomes for the actions to be learned. As discussed above, such knowledge-based intrinsic motivations might be represented by *unexpected changes*, as seems to be the case for the signals that are triggered by the superior colliculus, or by the *novelty signals* that seem to be produced by the hippocampus. There are a very few models in the literature of these kinds of mechanisms (e.g. Fiore et al., 2008; Sirois and Mareschal, 2004). In the most recent of them, we have shown how intrinsic reinforcements due to unexpected events might be particularly appropriate for driving the accumulation of skills that can be then deployed for maximizing extrinsic rewards (Mirolli et al., tted; Santucci et al., 2010). However, reinforcement signals that are so simple can only represent the very basic ingredients of intrinsically motivated learning. Complex actions can be acquired only on the basis of more complex learning signals, which we have argued might be represented by the dopaminergic signals triggered by the hippocampus on the basis of the perception of novel stimuli or of novel associations between stimuli. Modeling these more complex forms of intrinsically motivated learning is a completely open challenge.

A second fundamental issue for future research is checking whether our hypothesis about competence-based intrinsic motivations in real animals can be supported by animal research, and, in case it is, trying to understand which are the brain mechanisms that implement competence based motivation signals. In fact, while the study of competence-based intrinsic motivations is a fundamental research topic in the psychological literature on motivations in humans (Csikszentmihalyi, 1991; De Charms, 1968; Ryan and Deci, 2000), the topic of intrinsic motivations is now far less hot within the field of animal psychology, and this literature has been generally focusing more on demonstrating the existence of knowledge-based intrinsic motivations like novelty, surprise, and curiosity (Butler, 1953; Kish, 1955; Reed et al., 1996). Analogously, while we have identified the putative neural bases of knowledge-based intrinsic motivations (as the activation of the dopaminergic system by the detection of unexpected changes from the superior colliculus and, possibly, by the detection of novelty from the hippocampus), the identification of the possible neural mechanisms of competence-based intrinsic motivation signals (if any) is much harder at the moment, and hence constitutes a very important challenge for future research.

A further open issue is the identification and development of *hierarchical/modular* architectures that can support the kind of intrinsically-motivated cumulative acquisition of skills that we have discussed so far (Baldassarre and Mirolli, 2010). Two key requirements for such architectures is that they must (a) work with *continuous* state and action spaces, and (b) be capable of *learning autonomously*, in particular on the basis of (intrinsic) reinforcement learning signals. In particular, the capacity of these architectures to learn autonomously and with continuous state and action spaces would make them relevant both for controlling robots and for modeling real organisms (in particular if based on neural-network components). Various hierarchical architectures have been proposed in the literature but the majority of them is either dependent on abstract/symbolic representations of space and/or actions (Dayan and Hinton, 1993; Dietterich, 2000; Parr and Russell, 1997; Singh, 1992; Wiering and Schmidhuber, 1997) (see Barto and Mahadevan, 2003 for a review), or is trained in a supervised fashion (e.g. Haruno et al., 2001). Although there are exceptions (Caligiore et al., 2010; Doya et al., 2002; Konidaris and Barto, 2009; Provost et al., 2006; Schembri et al., 2007c), much more work is needed in this area, given its importance for implementing intrinsically-motivated cumulative learning.

A last important open issue is related to what we might call *goal-driven learning*. With this expression we refer to the possibility that intrinsic motivations might let a system form *goals*, and that it is these goals that drive the subsequent acquisition of skills The computational literature on hierarchical reinforcement learning has proposed several ways of creating goals, mostly as sub-goals derived from the final goal related to the task at hand (but see Jonsson and Barto, 2006; Vigorito and Barto, 2010 for exceptions). Here we refer instead to a situation in which there is no task to accomplish, but the system is endowed with intrinsic motivations that allow it to autonomously form goals with respect to which to acquire a competence, that is the capacity to pursue them in an efficient way. This capacity might be important for at least two reasons. First, the representation of the goal could allow the system to generate an intrinsic reward when the system succeeds in producing the action or the combination of actions that allow to achieve the goal. This might be crucial for allowing the system to learn the action or the action combination, i.e. to acquire the competence which reliably leads to the achievement of the goal itself. Second, it might make the system activate a number of perceptual and motor processes focused on the acquisition of a specific competence, while temporarily ignoring other stimuli which might generate other interfering intrinsic signals. For example, the system might focus attention on the portion of space and objects relevant for achieving the goal under focus, or might produce only those actions which are related to achieving goals similar to the pursued one. Eve this focusing mechanism might be crucial for learning complex skills. The importance of goals is evident, for example, in the IMRL systems proposed by Barto Barto et al. (2004), where options are created for leading the system to salient events (goals), but in that case goals are decided by the experimented and handcoded. Developing models in which goals are autonomously formed by the agent on the basis of intrinsic motivations and drive skill acquisition is another fundamental challenge for future research.

Aknowledgments

Thanks to Pierre-Yves Oudeyer, Andrew Barto, Kevin Gurney, and Jochen Triesh for their useful comments that substantially helped to improve the paper. Any remaining omission or mistake is our own blame.

Bibliography

Alexander, G., DeLong, M., and Strick, P. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Reviews Neuroscience*, 9:357–381.

Baldassarre, G. (2002). A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. *Journal of Cognitive Systems Research*, 3:5–13.

Baldassarre, G. and Mirolli, M. (2010). What are the key open challenges for understanding autonomous cumulative learning of skills? *AMD Newsletter*, 7(2):2–3.

Baldassarre, G. and Mirolli, M. (2012). Temporal difference competence based intrinsic motivation (TD-CB-IM): Td-error as an intrinsic reinforcement for deciding which skill to learn when. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Baranes, A. and Oudeyer, P.-Y. (2009). R-iac: Robust intrinsically motivated exploration and active learning. *IEEE Transactions on Autonomous Mental Development*, 1(3):155–169.

Baranes, A. and Oudeyer, P.-Y. (2010). Intrinsically motivated goal exploration for active motor learning in robots: A case study. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS 2010)*.

Barto, A. (1995). Adaptive critics and the basal ganglia. In Houk, J.C. Davis, J. and Beiser, D., editors, *Models of Information Processing in the Basal Ganglia*, pages 215–232. MIT Press, Cambridge, MA.

Barto, A., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *International Conference on Developmental Learning (ICDL)*, La Jolla, CA.

Barto, A., Sutton, R., and Anderson, C. (1983). Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics*, 13:834–846.

Barto, A. G. (2010). What are intrinsic reward signals? *AMD Newsletter*, 7(2):3.

Barto, A. G. (2012). Intrinsic motivation and reinforcement learning. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Barto, A. G. and Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, 13(4):341–379.

Berlyne, D. E. (1960). *Conflict, Arousal. and Curiosity.* McGraw-Hill, New York.

Botvinick, M. and Plaut, D. (2004). Doing without schema hierarchies: A recurrent connectionist approach to routine sequential action and its pathologies. *Psychological Review*, 111:395–429.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence Journal*, 47:139–159.

Butler, R. A. (1953). Discrimination learning by rhesus monkeys to visual-exploration motivation. *Journal of Comparative and Physiological Psychology*, 46(2):95–98.

Caligiore, D., Mirolli, M., Parisi, D., and Baldassarre, G. (2010). A bioinspired hierarchical reinforcement learning architecture for modeling learning of multiple skills with continuous states and actions. In *Proceedings of the Tenth International Conference on Epigenetic Robotics*, volume 149, Lund. Lund University Cognitive Studies.

Clark, A. (1997). *Being There: putting brain, body and world together again.* Oxford University Press, Oxford.

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience.* Harper Perennial, New York, NY.

Dalley, J. W., Cardinal, R. N., and Robbins, T. W. (2004). Prefrontal executive and cognitive functions in rodents: neural and neurochemical substrates. *Neuroscience and Biobehavioral Reviews*, 28(7):771–784.

Dayan, P. and Hinton, G. E. (1993). Feudal reinforcement learning. In *Advances in Neural Information Processing Systems 5*, pages 271–278. Morgan Kaufmann.

De Charms, R. (1968). *Personal Causation: The Internal Affective Determinants of Behavior.* Academic Press, New York, NY.

Deci, E. (1975). *Intrinsic motivation.* Plenum, New York.

Deci, E. L. and Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* Plenum, New York.

Dember, W. and Earl, R. (1957). Analysis of exploratory, manipulatory and curiosity behaviors. *Psychological Review*, 64:91–96.

Dietterich, T. (2000). Hierarchical reinforcement learning with the maxq value function decomposition. *Journal of Artificial Intelligence Research*, 13:227–303.

Dommett, E., Coizet, V., Blaha, C. D., Martindale, J., Lefebvre, V., Walton, N., Mayhew, J. E. W., Overton, P. G., and Redgrave, P. (2005). How visual stimuli activate dopaminergic neurons at short latency. *Science*, 307(5714):1476–1479.

Doya, K. (2000). Complementary roles of basal ganglia and cerebellum in learning and motor control. *Curr Opin Neurobiol*, 10(6):732–739.

Doya, K., Samejima, K., Katagiri, K.-i., and Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation*, 14(6):1347–1369.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, Stanford.

Fiore, V., Mannella, F., Mirolli, M., Gurney, K., and Baldassarre, G. (2008). Instrumental conditioning driven by neutral stimuli: A model tested with a simulated robotic rat. In *Proceedings of the Eight International Conference on Epigenetic Robotics*, number 139, pages 13–20, Lund. Lund University Cognitive Studies.

Fuster, J. (2001). The prefrontal cortex-an update: Time is of the essence. *Neuron*, 2:319–333.

Geisler, S., Derst, C., Veh, R. W., and Zahm, D. S. (2007). Glutamatergic afferents of the ventral tegmental area in the rat. *Journal of Neuroscience*, 27(21):5730–5743.

Grafton, S. T. and Hamilton, A. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Brain Mapping Movement Science*, 26(4):590–616.

Graziano, M. (2006). The organization of behavioral repertoire in motor cortex. *Annu Rev Neurosci*, 29:105–134.

Gurney, K., Lepora, N., Shah, A., Koene, A., and Redgrave, P. (2012). Action discovery and intrinsic motivation: a biologically constrained formalisation. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Gurney, K., Prescott, T. J., and Redgrave, P. (2001). A computational model of action selection in the basal ganglia I. A new functional anatomy. *Biological Cybernetics*, 84(6):401–410.

Harlow, H. F. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *Journal of Comparative and Physiological Psychology*, 43:289–294.

Harlow, H. F., Harlow, M. K., and Meyer, D. R. (1950). Learning motivated by a manipulation drive. *Journal of Experimental Psychology*, 40:228–234.

Haruno, M., Wolpert, D., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput*, 13:2201–2220.

Hebb, D. (1955). Drives and the conceptual nervous system. *Psychological Review*, 62:243–254.

Heidbreder, C. A. and Groenewegen, H. J. (2003). The medial prefrontal cortex in the rat: evidence for a dorso-ventral distinction based upon functional and anatomical characteristics. *Neuroscience and Biobehavioral Reviews*, 27(6):555–579.

Hof, P. M., Scherer, C., and Heuberger, P. S., editors (2009). *Model-Based Control: Bridging Rigorous Theory and Advanced Technology*. Springer, Berlin.

Horvitz, J. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, 96(4):651–656.

Houk, J., Adams, J., and Barto, A. (1995). A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In Houk, J., Davis, J., and Beiser, D., editors, *Models of Information Processing in the Basal Ganglia*, pages 249–270. MIT Press, Cambridge, MA.

Huang, X. and Weng, J. (2002). Novelty and reinforcement learning in the value system of developmental robots. In *Proceedings Second International Workshop on Epigenetic Robotics*, pages 47–55, Edinburgh, Scotland.

Hull, C. L. (1943). *Principles of behavior*. Appleton-century-crofts.

Hunt, H. (1965). Intrinsic motivation and its role in psychological development. *Nebraska Symposium on Motivation*, 13:189–282.

Joel, D., Niv, Y., and Ruppin, E. (2002). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw.*, 15(4):535–547.

Joel, D. and Weiner, I. (1994). The organization of the basal ganglia-thalamocortical circuits: Open interconnected rather than closed segregated. *Neuroscience*, 63(2):363 – 379.

Jonsson, A. and Barto, A. (2006). Causal graph based decomposition of factored mdps. *Journal of Machine Learning Research*, 7:2259–2301.

Jordan, M. I. and Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16:307–354.

Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, 22:51–66.

Kish, G. (1955). Learning when the onset of illumination is used as the reinforcing stimulus. *Journal of Comparative and Physiological Psychology*, 48(4):261–264.

Kish, G. and Antonitis, J. (1956). Unconditioned operant behavior in two homozygous strains of mice. *Journal of Genetic Psychology and Aging*, 88(1):121–129.

Konidaris, G. D. and Barto, A. G. (2009). Skill discovery in continuous reinforcement learning domains using skill chaining. In *Advances in Neural Information Processing Systems (NIPS 2009)*, pages 1015–1023.

Langton, C. G., editor (1989). *Artificial Life: The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems*, Redwood City, CA. Addison-Wesley.

Lee, R., Walker, R., Meeden, L., and Marshall, J. (2009). Category-based intrinsic motivation. In *Proceedings of the Ninth International Conference on Epigenetic Robotics*, volume 146, pages 81–88, Lund. Lund University Cognitive Studies.

Lisman, J. E. and Grace, A. A. (2005). The hippocampal-vta loop: controlling the entry of information into long-term memory. *Neuron*, 46(5):703–713.

Marshall, J., Blank, D., and Meeden, L. (2004). An emergent framework for self-motivation in developmental robotics. In *Proceedings of the Third International Conference on Development and Learning (ICDL 2004)*, pages 104–111, La Jolla, CA.

Merrick, K. and Maher, M. L. (2009). Motivated learning from interesting events: Adaptive, multitask learning agents for complex environments. *Adaptive Behavior*, 17(1):7–27.

Meyer, J.-A. and Wilson, S. W., editors (1990). *From animals to animats: Proceedings of the first international conference on simulation of adaptive behavior*, Cambridge, MA, USA. MIT Press.

Miller, E. and Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews Neuroscience*, 24:167–202.

Mink, J. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, 50(4):381–425.

Mirolli, M., Santucci, V. G., and Baldassarre, G. (Submitted). Phasic dopamine as a prediction error of intrinsic and extrinsic reinforcements driving both action acquisition and reward maximization: A simulated robotic study. *Neural Networks*.

Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, New York.

Montgomery, K. (1954). The role of exploratory drive in learning. *Journal of Comparative and Physiological Psychology*, 47:60–64.

Otmakova, N., Duzel, E., Deutch, A. Y., and Lisman, J. E. (2012). The hippocampal-vta loop: the role of novelty and motivation in controlling the entry of information into long-term memory. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Oudeyer, P.-Y. and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286.

Parr, R. and Russell, S. J. (1997). Reinforcement learning with hierarchies of machines. In *Advances in Neural Information Processing Systems*, Cambridge: MA. Mit Press.

Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT Press, Cambridge, MA.

Provost, J., Kuipers, B. J., and Miikkulainen, R. (2006). Developing navigation behavior through self-organizing distinctive state abstraction. *Connection Science*, 18(2):159–172.

Redgrave, P. (2007). Basal ganglia. *Scholarpedia*, 2(6):1825.

Redgrave, P. and Gurney, K. (2006). The short-latency dopamine signal: a role in discovering novel actions? *Nature Reviews Neuroscience*, 7(12):967–975.

Redgrave, P., Gurney, K., Stafford, T., Thirkettle, M., and Lewis, J. (2012). The role of the basal ganglia in discovering novel actions. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Redgrave, P., Prescott, T., and Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience*, 89:1009–1023.

Reed, P., Mitchell, C., and Nokes, T. (1996). Intrinsic reinforcing properties of putatively neutral stimuli in an instrumental two-lever discrimination task. *Animal Learning and Behavior*, 24:38–45.

Reynolds, J. N. and Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw*, 15(4-6):507–521.

Rizzolatti, G. and Luppino, G. (2001). The cortical motor system. *Neuron*, 31(6):889–901.

Romanelli, P., Esposito, V., Schaal, D. W., and Heit, G. (2005). Somatotopy in the basal ganglia: experimental and clinical evidence for segregated sensorimotor channels. *Brain research reviews*, 48:112–28.

Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ. Second edition.

Ryan, R. and Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25:54–67.

Santucci, V., Baldassarre, G., and Mirolli, M. (2010). Biological cumulative learning requires intrinsic motivations: a simulated robotic study on the development of visually-guided reaching. In *Proceedings of the Tenth International Conference on Epigenetic Robotics*, volume 149, Lund. Lund University Cognitive Studies.

Saunders, R. and Gero, J. (2002). Curious agents and situated design evaluations. In Gero, J. and Brazier, F., editors, *Agents in Design 2002*, pages 133–149, Sydney. Key Centre of Design Computing and Cognition, University of Sydney.

Schembri, M., Mirolli, M., and Baldassarre, G. (2007a). Evolution and learning in an intrinsically motivated reinforcement learning robot. In *Advances in Artificial Life. Proceedings of the 9th European Conference on Artificial Life*, volume 4648 of *LNAI*, pages 294–333, Berlin. Springer.

Schembri, M., Mirolli, M., and Baldassarre, G. (2007b). Evolving childhood's length and learning parameters in an intrinsically motivated reinforcement learning robot. In *Proceedings of the Seventh International Conference on Epigenetic Robotics*, pages 141–148, Lund. Lund University Cognitive Studies.

Schembri, M., Mirolli, M., and Baldassarre, G. (2007c). Evolving internal reinforcers for an intrinsically motivated reinforcement-learning robot. In *Proceedings of the 6th International Conference on Development and Learning*, pages E1–6, London. Imperial College.

Schmidhuber, J. (1991a). Curious model-building control systems. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2, pages 1458–1463, Singapore. IEEE.

Schmidhuber, J. (1991b). A possibility for implementing curiosity and boredom in model-building neural controllers. In *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*, pages 222–227, Cambridge, MA. MIT Press.

Schmidhuber, J. (2002). Exploring the predictable. In Ghosh, S. and Tsutsui, T., editors, *Advances in Evolutionary Computing*, pages 579–612. Springer-Verlag, Berlin.

Schmidhuber, J. (2012). Maximizing fun by creating data with easily reducible subjective complexity. In Baldassarre, G. and Mirolli, M., editors, *Intrinsically Motivated Learning in Natural and Artificial Systems*. Springer-Verlag. (this volume).

Schultz, W., Dayan, P., and Montague, P. (1997). A neural substrate of prediction and reward. *Science*, 275:1593–1599.

Schultz, W. and Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Reviews Neuroscience*, 23:473–500.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press, Cambridge, MA. 3rd Edition.

Singh, S. P. (1992). Transfer of learning by composing solutions of elemental sequential tasks. *Machine Learning*, 8:323–339.

Sirois, S. and Mareschal, D. (2004). An interacting systems model of infant habituation. *Journal of Cognitive Neuroscience*, 16(8):1352–1362.

Storck, J., Hochreiter, S., and Schmidhuber, J. (1995). Reinforcement-driven information acquisition in non-deterministic environments. In *Proceedings of*

ICANN'95, volume 2, pages 159–164, Paris.

Stout, A. and Barto, A. G. (2010). Competence progress intrinsic motivation. In *Proceedings of the 9th International Conference on Development and Learning (ICDL 2010)*, pages 257–262.

Stout, A., Konidaris, G. D., and Barto, A. G. (2005). Intrinsically motivated reinforcement learning: A promising framework for developmental robot learning. In *Proceedings of the AAAI Spring Symposium on Developmental Robotics*, Stanford, CA.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA.

Sutton, R., Precup, D., and Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44.

Taber, M., Das, S., and Fibiger, H. (1995). Cortical regulation of subcortical dopamine release: mediation via the ventral tegmental area. *Journal of Neurochemistry*, 65(3):1407–1410.

Tani, J., Nishimoto, R., and Paine, R. (2008). Achieving 'organic compositionality' through self-organization: Reviews on brain-inspired robotics experiments. *Neural Networks*, 21:584–603.

Tani, J. and Nolfi, S. (1999). Learning to perceive the world as articulated: an approach for hierarchical learning in sensory-motor systems. *Neural Networks*, 12:1131–1141.

Vigorito, C. and Barto, A. (2010). Intrinsically motivated hierarchical skill learning in structured environments. *IEEE Transactions on Autonomous Mental Development*, 2(2):83–90.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291:599–600.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66:297–333.

Wiering, M. and Schmidhuber, J. (1997). Hq-learning. *Adaptive Behavior*, 6:219–246.

Yamashita, Y. and Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. *PLoS Computational Biology*, 4(11):e1000220.

Yin, H. H. and Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7:464–476.