How can we define intrinsic motivation ?

Pierre-Yves Oudeyer, Frédéric Kaplan

How can we define intrinsic motivation?*

Pierre-Yves Oudeyer [1], Frederic Kaplan [2]
[1]**INRIA Bordeaux Sud-Ouest**
351, Cours de la Libration, 33405 Talence, France
[2]**Ecole Polytechnique Federale de Lausanne**
EPFL - CRAFT, CE 1 628, Station 1, CH- 1015 Lausanne, Switzerland

Abstract

Intrinsic motivation is a crucial mechanism for open-ended cognitive development since it is the driver of spontaneous exploration and curiosity. Yet, it has so far only been conceptualized in ad hoc manners in the epigenetic robotics community. After reviewing different approaches to intrinsic motivation in psychology, this paper presents a unified definition of intrinsic motivation, based on the theory of Daniel Berlyne. Based on this definition, we propose a landscape of types of computational approaches, making it possible to position existing and future models relative to each other, and we show that important approaches are still to be explored.

1. Introduction

Intrinsic motivation has been a topic of growing interest in the developmental robotics and reinforcement learning communities in the recent years (Barto et al., 2004, Oudeyer et al., 2007). This concept comes from psychology, and has been argued to be crucial for open-ended cognitive development (Ryan and Deci, 2000). In particular, psychologists have proposed that it is the mechanism that explains the spontaneous exploratory behaviors observed in humans, and infants in particular (Berlyne, 1965). Researchers in developmental robotics and reinforcement learning have proposed that intrinsic motivation might allow the acquisition of general and re-usable skills (Barto et al., 2004), increase the efficiency of learning when considered as an active learning mechanism (Thrun, 1995), guide and structure exploration in large spaces (Oudeyer et al., 2007). Yet, intrinsic motivation has been conceptualized rather differently by computer scientists, through the implementation of a number of ad hoc models, e.g. (Schmidhuber, 1991, Thrun, 1995, Huang and Weng, 2002, Kaplan and Oudeyer, 2003, Marshall et al., 2004, Andry et al., 2004, Oudeyer and Kaplan, 2006, Barto et al., 2004, Bonarini et al., 2006, Oudeyer et al., 2007, Schembri et al., 2007, Merrick, 2008). For example, intrinsic motivation has sometimes been confused with internal motivations. In fact, a unified definition does not seem to exist yet, and no framework exists that allows to relate easily different intrinsic motivation mechanisms to each others. We propose in this paper to come back thoroughly on the ways intrinsic motivation has been approached in psychology, in order to extract a principled definition. Based on this principled definition, we will sketch a landscape of possible computational approaches to intrinsic motivation. This will enable us to show that most of this landscape is still unexplored, and opens avenues for research in the future.

*This paper is based on (Oudeyer and Kaplan, 2007), but presents a novel proposition to use Berlyne's concept of "collative variable" to define intrinsic motivation.

2. What is intrinsic motivation? The psychologists' point of view

2.1 Activities pursued for their own sake

According to (Ryan and Deci, 2000) (pp. 56),

> *Intrinsic motivation is defined as the doing of an activity for its inherent satisfaction rather than for some separable consequence. When intrinsically motivated, a person is moved to act for the fun or challenge entailed rather than because of external products, pressures or reward.*

Intrinsic motivation is clearly visible in young infants, that consistently try to grasp, throw, bite, squash or shout at new objects they encounter. Even if less important as they grow, human adults are still often intrinsically motivated while they play crosswords, make paintings, do gardening or just read novels or watch movies. Yet, to get a clearer picture of intrinsic motivation, one needs to understand that it has been defined by contrast to extrinsic motivation:

> *Extrinsic motivation is a construct that pertains whenever an activity is done in order to attain some separable outcome. Extrinsic motivation thus contrasts with intrinsic motivation, which refers to doing an activity simply for the enjoyment of the activity itself, rather than its instrumental value.* (Ryan and Deci, 2000)

Intrinsic is not a synonym of internal. We see that, according to this approach, a central feature that differentiates intrinsic and extrinsic motivation is *instrumentalization.* We also see that the concepts of intrinsic and extrinsic motivations form a different distinction than the one between internal and external motivations which is sometimes made in the cognitive robotics and reinforcement learning literature. Moreover, in this computational literature, "intrinsic" is sometimes used as a synonym to "internal", and "extrinsic" as a synonym to "external". Yet, it is in fact a confusion. Internal motivations involve reward that are produced within the organism, whichever they are, and external motivations involve rewards that are produced outside the organism (e.g. coming from social partners). The intrinsic/extrinsic distinction, on the contrary, is not a distinction based on the location of origin of the reward, but on the kind of reward, as it will become more clear below with the approach proposed by Berlyne.

Let us give examples to be more clear. For example, a child that does thoroughly his homework might be motivated by avoiding the sanctions that his parents could give him in case he would not do it. The cause for action is here clearly external, and the homework is not done for its own sake but for the separate outcome of not getting sanctions. Here the child is extrinsically and externally motivated.

On the other hand, it is possible that a child could do thoroughly his homework because he is persuaded that it will help him get the job he dreams of, later when he will be an adult. In this case, the cause for action is internally generated, and the homework is again not achieved for its own sake but because the child thinks it will lead to the separate outcome of getting a good job. Here the child is internally and extrinsically motivated.

Finally, it is also possible that a child does thoroughly his homework for the fun of it, and because he experiences pleasure in the discovery of new knowledge or considers for example its math problem just as fun as playing a video game. In this case, his behavior is intrinsically and internally motivated.

These different kinds of motivations can also sometimes be superposed or interleaved in the same global activity. For example, it is quite possible that a child doing his homework is partly extrinsically motivated by getting a high grade at the exam and partly intrinsically motivated by learning new interesting things. Thus, the same activity can be at the same time intrinsically *and* extrinsically motivating. Also, for example, imagine a child that is intrinsically motivated by playing tennis but has to ride its bicycle to get to the tennis court (and does not like particularly riding bicycles). In this case, the riding of the bicycle is an internal and extrinsically motivated behavior that spins out of the intrinsically motivated behavior of playing tennis.

2.2 *What makes an activity intrinsically motivating?*

Given this broad distinction between intrinsic and extrinsic motivation, psychologists have tried to build theories about which features of activities make them intrinsically motivating for some people (and not all) at some times (the same activity might be intrinsically motivating for a person at a given time, but no more later on). They have studied how these motivations could be functionally implemented in an organism, humans in particular, and several theoretical directions have been presented.

Drives to manipulate, drives to explore In the 1950s, psychologists started by trying to give an account of intrinsic motivation and exploratory activities on the basis of the theory of drives (Hull, 1943), which are specific tissue deficits like hunger or pain that the organisms try to reduce. For example, (Montgomery, 1954) proposed a drive for exploration and (Harlow, 1950) a drive to manipulate. This drive naming approach had many short-comings which were criticized in detail by White in 1959 (White, 1959): intrinsically motivated exploratory activities have a fundamentally different dynamics. Indeed, they are not homeostatic: the general tendency to explore is not a consummatory response to a stressful perturbation of the organism's body.

Reduction of cognitive dissonance Some researchers then proposed another conceptualization. Festinger's theory of cognitive dissonance (Festinger, 1957) asserted that organisms are motivated to reduce dissonance, which is the incompatibility between internal cognitive structures and the situations currently perceived. Fifteen years later a related view was articulated by Kagan stating that a primary motivation for humans is the reduction of uncertainty in the sense of the "incompatibility between (two or more) cognitive structures, between cognitive structure and experience, or between structures and behavior" (Kagan, 1972). However, these theories were criticized on the basis that much human behavior is also intended to *increase* uncertainty, and not only to reduce it. Human seem to look for some forms of optimality between completely uncertain and completely certain situations.

Optimal incongruity In 1965, Hunt developed the idea that children and adult look for optimal incongruity (Hunt, 1965). He regarded children as information-processing systems and stated that interesting stimuli were those where there was a discrepancy between the perceived and standard levels of the stimuli. For Dember and Earl, the incongruity or discrepancy in intrinsically-motivated behaviors was between a person's expectations and the properties of the stimulus (Dember and Earl, 1957). Berlyne developed similar notions as he observed that the most rewarding situations were those with an intermediate level of novelty, between already familiar and completely new situations (Berlyne, 1960).

Motivation for effectance, personal causation, competence and self-determination Eventually, a last group of researchers preferred the concept of challenge to the notion of optimal incongruity. These researchers stated that what was driving human behavior was a motivation for effectance (White, 1959), personal causation (De Charms, 1968), competence and self-determination (Deci and Ryan, 1985). Basically, these approaches argue that what motivates people is the degree of control they can have on other people, external objects and themselves, or in other words, the amount of effective interaction. In an analogous manner, the concept of optimal challenge has been put forward, such as for example in the theory of "Flow" (Csikszentmihalyi, 1991).

2.3 Collative variables

These diverse theoretical approaches to intrinsic motivation and to the properties that shall make certain activities intrinsically interesting/motivating have been proposed and published by diverse research communities within psychology, in such a way that still today there is no consensus among these communities on a unified or integrated view of intrinsic motivation. Even more, it could be argued that distinguishing intrinsic and extrinsic motivation based on instrumentalization can be circular (Oudeyer and Kaplan, 2007). Yet, a convincing integrated non-circular view has actually been proposed in the 60's by Daniel Berlyne (Berlyne, 1965), and shall be used as a fruitful theoretical reference for developmental roboticists. The central concept of this integrated approach to intrinsic motivation is that of "collative variables", as explained in the following quotations:

> *The probability and direction of specific exploratory responses can apparently be influenced by many properties of external stimulation, as well as by many intraorganism variables. They can, no doubt, be influenced by stimulus intensity, color, pitch, and association with biological gratification and punishment, ... [but] the paramount determinants of specific exploration are, however, a group of stimulus properties to which we commonly refer by such words as "novelty", "change", "surprisingness", "incongruity", "complexity", "ambiguity", and "indistinctiveness".* (Berlyne, 1965), pp. 245.

> *... these properties possess close links with the concepts of information theory, and they can, in fact, all be discussed in information-theoretic terminology. In the case of "ambiguity" and "indistinctiveness", there is uncertainty due to a gap in available information. In some forms of "novelty" and "complexity", there is uncertainty about how a pattern should be categorized, that is, what labeling responses should be attached to it and what overt response is appropriate to it. When one portion of a "complex" pattern or of a sequence of "novel" stimuli is perceived, there is uncertainty about what will be perceived next. In the case of "surprisingness" and "incongruity", there is discrepancy between information embodied in expectations and information embodied in what is perceived. For these reasons, the term "collative" is proposed as an epithet to denote all these stimulus properties collectively, since they all depend on collation or comparison of information from different stimulus elements, whether they be elements belonging to the present, past or elements that are simultaneously present in different parts of one stimulus field".*
> *It should be pointed out that the uncertainty we are discussing here is "subjective uncertainty", which is a function of subjective probabilities, analogous to the "objective" uncertainty (that is, the standard information-theoretic concept of uncertainty) that is a function of objective probabilities.* (Berlyne, 1965), pp. 245-246.

> *Drive increases when an organism is subjected to a physiological disturbance, such as those accompanying hunger, thirst, and sexual excitement, or to noxious external agents. It can also be raised by stimuli, external or internal, that have been regularly paired with such disturbances. These motivating conditions, which involve organs other than the sense organs and the nervous system, may be called "sources of extrinsic motivation". They can undoubtedly actuate exploratory or epistemic behavior, as when as a person seeks information for the solution of a practical problem or for the social status that erudition will bring him. There are, however, other, "intrinsic" forms of motivation which collaborate with extrinsic motivation in regulating exploratory or epistemic activity but are also capable of actuating exploratory or epistemic activity on their own. Intrinsic motivation depends primarily on the collative properties of the external environment.* (Berlyne, 1965), pp. 252.

This leads us to the following characterization of intrinsic motivation:

An activity or an experienced situation, be it physical or imaginary, is intrinsically motivating for an autonomous entity if its interest depends primarily on the collation or comparison of *information* from different stimuli and *independently of their semantics*, whether they be physical or imaginary stimuli (i.e. measured by physical sensors or by internal "software" sensors) perceived in the present or in the past (in which case they will typically be internally represented and compressed by the brain through learning) or stimuli that are simultaneously present in different parts of one stimulus field.

Most importantly, the information that is compared has to be understood in an information theoretic perspective, in which what is considered is the intrinsic mathematical structure of the values of stimuli, independently of their meaning. As a consequence, measures which pre-suppose the meaning of stimuli, i.e. the meaning of sensorimotor channels (e.g. the fact that a measure is a measure of energy or temperature or color), do not characterize intrinsically motivating activities or situations.

In practice, this means that a typical intrinsic motivation is for example a motivation to search for surprising situations, whatever they are, and that a typical non-intrinsic motivation is for example a motivation to search for food or water in order to maintain the internal metabolic equilibrium of the body.

3. Carving the landscape of computational implementations of intrinsic motivation

The characterization of intrinsic motivation based on Berlyne's theory can be used as a common conceptual framework to interprete and compare the various computational architectures that have been proposed in the literature for implementing forms of intrinsic motivation, e.g. (Schmidhuber, 1991, Thrun, 1995, Huang and Weng, 2002, Kaplan and Oudeyer, 2003, Marshall et al., 2004, Oudeyer and Kaplan, 2006, Sporns and Lungarella, 2006, Barto and Simsek, 2005, Schembri et al., 2007, Oudeyer et al., 2007, Capdepuy et al., 2007, Merrick, 2008). Moreover, while this characterization excludes a number of internal motivation mechanisms from intrinsic motivation mechanisms, its generic formulation also encompasses many mechanisms that have not been explored yet in the computational literature.

The goal of this section is to present a landscape of potential computational implementations of intrinsic motivation, allowing us to set the basis of a typological and formal framework that may allow researchers to understand better and map the space of possible models. The length of this article does not allow us to present exhaustively this landscape, and thus we will focus on representative types of mechanisms.

This landscape consists in presenting and organizing a set of measures, based on collative variables, that may be used by an autonomous entity to evaluate the intrinsic interestingness of an activity or a situation. One possible cognitive architecture in which these measures could be embedded is that of computational reinforcement learning (Sutton and Barto, 1998). In this case, these measures correspond to internally generated rewards that an action selection system based on algorithms such as Q-learning or Sarsa shall use as input (possibly together with other sources of internal or external rewards) in order to select actions that will maximize the expected cumulated sum of these rewards obtained in the future. For this reasons, measures of interestingness of a situation or activity e^k at a given time t can be considered as computational definitions of internal rewards generated upon the encountering of e^k at time t, and will be denoted $r(e^k, t)$.

Some of the types of mechanisms that will be presented have already been implemented in the literature, in which case we will provide references to such implementations, and others have not been explored yet. In all cases, the objective of this list is to show the variety and interrelations of potential mechanisms, but not to comment on the kind of behavior that might result from using these measures of interestingness in an autonomous entity, which has partially been done in other papers and partially will have to be done in future research.

In the following, we organize the space of computational models of intrinsic motivation into three broad classes that all share the same formal notion of a sensorimotor flow experienced by a robot. We assume that the typical robot is characterized by a number of sensory channels, denoted s_i, and motor channels denoted m_i, whose values continuously flow with time, hence the notations $s_i(t)$ and $m_i(t)$. The vector of all sensorimotor values at time t is denoted $SM(t)$. Three features are important for the following computational models:

1. these channels may correspond to any kind of physical or internal variable of a robot, which can be low-level (e.g. the color values of the pixels of a camera, the instantaneous intensity of sound perceived by a microphone, the value of a motor joint, ...) or higher-level (e.g. the presence or absence of a face in an image or its position, the identity of a person speaking, the triggering of a whole grasping movement, ...);
2. what these sensory channels actually are, i.e. their "meaning", is NOT taken into account;
3. the set of sensorimotor channels taken into account in intrinsic motivation measures of a situation may be smaller than the set of all sensorimotor channels available to the robot.

We will now present three broad types of measures of interestingness that can characterize intrinsic motivation and are based on collative variables:

1. Knowledge based models, in which interestingness is related to comparisons between the predicted flow of sensorimotor values, based on an internal forward model, with the actual flow of values; This typically leads to adaptive motivation if the model is learnt (adaptive motivation refers to mechanisms that assign different levels of interest to the same situation/activity depending on the particular moment in development where it is encountered).

2. Competence based models, in which interestingness is related to comparisons between self-generated goals, which are particular configurations in the sensorimotor space, and the extent to which they are reached in practice, based on an internal inverse model that may be learnt. Thus, these comparisons characterize the degree of performance/competence of an agent and also typically lead to adaptive motivation if the model is learnt.

3. Morphological models, in which interestingness is related to measures of the immediate structural relationships among multiple sensorimotor channels which are not based on long-term knowledge or competence previously acquired by the agent. This typically leads to fixed motivation.

3.1 Knowledge-based models of intrinsic motivation

A first computational approach to intrinsic motivation is based on measures of dissonances (or resonances) between the situations experienced by a robot and the knowledge and expectations that the robot has about these situations. Here the word "situation" might refer as well to a passive observation activity in which a robot does nothing but focus its attention on a particular aspect of the environment, as to an active activity in which the robot performs actions and compares the actual outcome of its actions to its knowledge and expectations about these actions.

Within this approach, there are two sub-approaches related to the way knowledge and expectations are represented: information theoretic/distributional and predictive.

3.1.1 Information theoretic and distributional models

This approach is based on the use of representations, built by the robot, that estimate the distributions of probabilities of observing certain events e^k in particular contexts, defined as mathematical configurations in the sensorimotor flow. There are several types of such events, but the probabilities that are measured are typically either the probability of observing a certain state SM^k in the sensorimotor flow, denoted $P(SM^k)$, or the probability of observing particular transitions between states, such as $P(SM^k(t), SM^l(t+1))$, or the probability of observing a particular state after having observed a given state $P(SM^k(t+1)|SM^l(t))$. Here, the states SM^k can either be direct numerical prototypes or complete regions within the sensorimotor space (and it may involve a mechanism for discretizing the space). In the following, we will consider all these eventualities possible and just use the general notation $P(e^k)$. We will assume that the robot possesses a mechanism that allows it to build internally, and as it experiences the world, an estimation of the probability distribution of events across the whole space E of possible events (but the space of possible events is not predefined and should also be discovered by the robot, so typically this is an initially empty space that grows with experience). Finally, we use the concept of entropy, which characterizes the shape of the distribution function, for discretized spaces:

$$H(E) = -\sum_{e^k \in E} P(e^k) ln(P(e^k)) \tag{1}$$

Uncertainty motivation (UM) The tendency to be intrinsically attracted by novelty has often been used as an example in the literature on intrinsic motivation. A straightforward manner to computationally implement it is to build a system that, for every event e^k that is actually observed, will generate a reward $r(e^k)$ inversely proportional to its probability of observation:

$$r(e^k, t) = C \cdot (1 - P(e^k, t)) \tag{2}$$

where C is a constant. Various models based on UM-like mechanisms were implemented in the computational literature (e.g. (Huang and Weng, 2002))

Information gain motivation (IGM) It has also often been proposed in psychology and education that humans have a natural propensity to learn and assimilate (Ryan and Deci, 2000). In information theoretic terms, this notion of assimilation or of "pleasure of learning" can be modeled by the decrease of uncertainty in the knowledge that the robot has of the world after an event e^k has happened:

$$r(e^k, t) = C \cdot (H(E, t) - H(E, t+1)) \tag{3}$$

Examples of implementation of this information gain motivation can be found for instance in (Fedorov, 1972, Roy and McCallum, 2001) (but note that in these paper the term "motivation system" is not used). It should be noted that, in practice, it is not necessarily tractable in continuous spaces. Actually, this is potentially a common problem to all distributional approaches.

Empowerment (EM) Empowerment (Capdepuy et al., 2007) is a reward measure that pushes an agent to produce sequences of actions that can transfer a maximal amount of information to its sensors through the environment. It is defined as the channel capacity from the sequence of actions $A_t, A_{t+1}, ..., A_{t+n-1}$ to the perceptions S_{t+n} after an arbitrary number of timesteps:

$$r(A_t, A_{t+1}, ..., A_{t+n-1} \to S_{t+n}) =$$
$$max_{p(\vec{a})} I(A_t, A_{t+1}, ..., A_{t+n-1}, S_{t+n})$$

where $p(\vec{a})$ is the probability distribution function of the action sequences $\vec{a} = (a_t, a_{t+1}, ..., a_{t+n-1})$ and I is mutual information. (Capdepuy et al., 2007) has shown how it could foster the emergence of complex behavior.

3.1.2 Predictive models

Often, knowledge and expectations in robots are not represented by complete probability distributions, but rather based on the use of predictors such as neural networks or support vector machines that make direct predictions about future events . In this kind of architecture, it is also possible to define computationally various forms of intrinsic motivations. These predictors, denoted Π, are typically used to predict some properties Pr^k or sensorimotor states SM^k that will happen in the future (close or far) given the current sensorimotor context $SM(t)$ and possibly the past sensorimotor context. Similarly to above, we will denote all properties and states under the generic notation e^k. We will also use the notation $SM(\to t)$ to denote a structure which encodes the current sensorimotor context and possibly the past contexts. Thus, a general prediction of a system will be denoted:

$$\Pi(SM(\to t)) = \tilde{e}^k(t+1) \tag{4}$$

We then define $E_r(t)$ as the error of this prediction, being the distance between the predicted event $\tilde{e}^k(t+1)$ and the event that actually happens $e^k(t+1)$:

$$E_r(t) = \|\tilde{e}^k(t+1) - e^k(t+1)\| \tag{5}$$

Predictive novelty motivation (NM) It then comes naturally to propose a first manner to model a motivation for novelty in this framework. Interesting situations are those for which the prediction errors are highest:

$$r(SM(\rightarrow t)) = C \cdot E_r(t) \tag{6}$$

where C is a constant. Examples of implementation of this kind of motivation system can be found for example in (Thrun, 1995, Barto et al., 2004).

Intermediate level of novelty motivation (ILNM) According to psychologists that proposed that humans are attracted by situations of intermediate/optimal incongruity, one can update the previous mechanism by introducing a threshold E_r^σ that defines this intermediate level of novelty:

$$r(SM(\rightarrow t)) = C_1 \cdot e^{-C_2 \cdot \|E_r(t) - E_r^\sigma\|^2} \tag{7}$$

where C_1 and C_2 are constants. Yet, this definition has the drawback of leaving the tuning of the threshold to the intuition of the human engineer. As a matter of fact, having a single threshold for the whole sensorimotor space might even be quite problematic in practice, since notions of novelty and similarities might vary a lot in different parts of that space, and developing mechanisms for automatic adaptive thresholding is a difficult problem.

Learning progress motivation (LPM) Several researchers have proposed another manner to model optimal incongruity which avoids the problem of setting a threshold, and is related to the information gain measurement described in the information theoretic section above. It consists in modeling intrinsic motivation with a system that generates rewards when predictions improve over time. Thus, the system will try to maximize prediction progress, i.e. the decrease of prediction errors, i.e. effectively reward knowledge *acquisition* per se. This corresponds to the concept of epistemic curiosity proposed by Berlyne (Berlyne, 1965). A first computational formalization was proposed in (Schmidhuber, 1991). Another kind of computational formalization was proposed in (Oudeyer et al., 2007), where "prediction progress" was referred as "learning progress". To get a formal model, one needs to be precise and subtle in how the decrease is computed. Indeed, as argued in (Oudeyer et al., 2007), the possible naive implementation comparing prediction errors between a window around time t and a window around time $t - \theta$ is in fact nonsense: this may for example attribute a high reward to the transition between a situation in which a robot is trying to predict the movement of a leaf in the wind (very unpredictable) to a situation in which it just stares at a white wall trying to predict whether its color will change (very predictable). The system should not try to compare very different sensorimotor situations and qualitatively different predictions.

A first proposition to compute learning progress and get around this problem was proposed by (Schmidhuber, 1991). It consists in measuring the difference in prediction error of the predictor Π, about the same sensorimotor context $SM(\rightarrow t)$, between the first prediction and a second prediction made just after the predictor has been updated with a learning rule:

$$r(SM \rightarrow t) = E_r(t) - E_r^{'}(t) \tag{8}$$

where

$$E_r^{'}(t) = \|\Pi^{'}(SM(\rightarrow t)) - e^k(t+1)\| \tag{9}$$

with $\Pi^{'}$ being the updated predictor after the learning update due to the prediction $\Pi(SM(\rightarrow t))$ and the perception of the actual consequence $e^k(t+1)$.

Another approach to compute learning progress, presented in (Oudeyer et al., 2007), is to use a mechanism that will allow the robot to group similar situations into regions $\mathcal{R}_n$ within which comparison is meaningful. The number and boundaries of these regions are typically adaptively updated (Oudeyer et al., 2007). Then, for each of these regions, the robot monitors the evolution of prediction errors, and makes a model of their global derivative in the past, which defines learning progress, and thus reward, in these regions. Mathematically:

$$r(SM(\rightarrow t)) =< E_r^{\mathcal{R}_n}(t-\theta) > - < E_r^{\mathcal{R}_n}(t) > \tag{10}$$

with $SM(t)$ belonging to region $\mathcal{R}_n$ and where $< E_r^{\mathcal{R}_n}(t) >$ is the mean of predictions errors made by the predictor in the last τ predictions made about sensorimotor situations $SM(t)$ belonging to region $\mathcal{R}_n$. A detailed study about how to implement such a system is provided in (Oudeyer et al., 2007).

Predictive familiarity motivation (FM) In the psychology literature, intrinsic motivations refer generally to mechanisms that push organisms to explore their environment. Yet, there are direct variants of previous computational systems that are both simple and correspond intuitively to existing forms of human motivation. For example, a slight mathematical variation of NM would model a motivation to search for situation which are very predictable, and thus familiar:

$$r(SM(\rightarrow t)) = \frac{C}{E_r(t)} \tag{11}$$

where C is a constant. It would actually be sound to consider this kind of motivation as intrinsic, in spite of the fact that it will typically not push an organism to explore its environment. FM is also related to cognitive homeostasis, and some experiments have shown its potential impact on development (Andry et al., 2004).

3.2 Competence-based models of intrinsic motivation

A second major computational approach to intrinsic motivation is based on measures of competence that an agent has for achieving self-determined results or goals. It is directly inspired from important psychological theories of effectance (White, 1959), personal causation (De Charms, 1968), competence and self-determination (Deci and Ryan, 1985), and "Flow" (Csikszentmihalyi, 1991). Central here is the concept of "challenge", with associated measures of difficulty as well as measures of actual performance. A "challenge" or "goal" here will be any sensorimotor configuration SM^k, or any set $\{P_k\}$ of properties of a sensorimotor configuration, that the individual sets by itself and that it tries to achieve through action. A challenge/goal is here self-determined, denoted $g^k = \{P_k\}$. It is the properties of the achievement process, rather than the "meaning" of the particular goal being achieved, that will determine the level of interestingness of the associated activity. While prediction mechanisms or probability models, as used in previous sections, can be used in the goal-reaching architecture, they are not mandatory (for example, one can implement systems that try to achieve self-generated goals through Q-learning and never explicitly make predictions of future sensorimotor contexts). Furthermore, while in some cases, certain competence-based and knowledge-based models of intrinsic motivation might be somewhat similar, they may often produce very different behaviors. Indeed, the capacity to predict what happens in a situation can be sometimes only loosely coupled to the capacity to modify a situation in order to achieve a given self-determined goal.

More technically, we will assume here a cognitive architecture in which there are two levels of action, associated to two time scales for decision. First, there is a high-level of action consisting in choosing what goals shall be explored for reaching (thus, this is not a physical but a mental action). This flow of choices is associated to a first time scale, in which the time when a goal is set is denoted t_g. Second, there is a lower-level of action consisting in choosing what to do in order to reach the goals. Whenever a goal $g^k(t_g)$ is set, there is a "know-how" module $KH(t_g)$ that is responsible for planning the lower-level actions in order to reach it and that learns through experience. After a certain amount of time, bounded for example by a timeout T_g, a motivation module compares the goal that was initially set and the current situation to assess to what extent it was reached, i.e. measure the competence of the agent on goal g^k at time t_g:

$$l_a(g_k, t_g) = \|\widetilde{g_k(t_g)} - g_k(t_g))\| \tag{12}$$

The "interestingness", and thus reward value, of the goal g^k is then derived from this competence measure. The next goal is then chosen at time $t_g + 1$ in such a way that the expected cumulated sum of these rewards in the future will be maximal, and traditional reinforcement learning algorithm can be used for this selection of action (i.e. for the selection of adequate goals).

Goal setting and reaching episodes are related to temporally extended actions in option theory (Sutton et al., 1999). However, to our knowledge, this paper presents competence-based models of intrinsic motivation that seem to have been very limitedly explored so far.

Maximizing incompetence motivation (IM) A first competence-based approach to intrinsic motivation can be a system which pushes the robot to set challenges/goals for which its performance is lowest. This is a motivation for maximally difficult challenges. This can be implemented as:

$$r(SM(\to t), g_k, t_g) = C \cdot l_a(g_k, t_g) \tag{13}$$

Note that here and everywhere in the competence based approaches, rewards are generated only at the end of episodes.

Maximizing competence progress - aka Flow motivation (CPM) Maximizing incompetence does not model very well the psychological models of optimal challenge and "flow" proposed by (Csikszentmihalyi, 1991). Flow refers to the state of pleasure related to activities for which difficulty is optimal: neither too easy nor too difficult. As difficulty of a goal can be modeled by the (mean) performance in achieving this goal, a possible manner to model flow would be to introduce two thresholds defining the zone of optimal difficulty. Yet, the use of thresholds can be rather fragile, require hand tuning and possibly complex adaptive mechanism to update these thresholds during the robot's lifetime. Another approach can be taken, which avoids the use of thresholds. It consists in defining the interestingness of a challenge as the competence progress that is experienced as the robot repeatedly tries to achieve it. So, a challenge for which a robot is bad initially but for which it is rapidly becoming good will be highly rewarding. Thus, a first manner to implement CPM would be:

$$r(SM(\to t), g_k, t_g) = C \cdot (l_a(g_k, t_g - \theta) - l_a(g_k, t_g)) \tag{14}$$

corresponding to the difference between the current performance for task g_k and the performance corresponding to the last time g_k was tried, at a time denoted $t_g - \theta$. Again, because of possible high variance in goal achievement, one could use smoothed differences:

$$r(SM(\to t), g_k, t_g) = C \cdot (< l_a(g_k, t_g - \theta) > - < l_a(g_k, t_g) >) \tag{15}$$

with $< l_a(g_k), t_g) >$ being the mean performance in trying to reach g_k in the last τ corresponding episodes, and $< l_a(g_k), t_g - \theta) >$ being the mean performance in trying to reach g_k between episodes $t_g - \theta - \tau$ and $t_g - \theta$.

3.3 Morphological models of intrinsic motivation

The two previous computational approaches to motivation were based on measures comparing information characterizing a stimulus perceived in the present and information characterizing stimuli perceived in the past and represented in memory. A third approach that can be taken is based on the comparison of information characterizing several pieces of stimuli perceived at the same time in several parts of the perceptive field. Pragmatically, this approach consists in attributing interest depending on morphological mathematical properties of the current flow of sensorimotor values, irrespective of what the internal cognitive system might predict or master.

Synchronicity motivation (SyncM) One typical example of this type of intrinsic motivation mechanism is based on synchronicity. The synchronicity motivation is based on an information theoretic measure of short-term correlation (or reduced information distance) between a number of sensorimotor channels. With such a motivation, situations for which there is a high short-term correlation between a maximally large number of sensorimotor channels are very interesting. This can be formalized in the following manner.

Let us consider that the sensorimotor space SM is a set of n information sources $\{SM_i\}$ and that possible values for these information sources typically correspond to elements belonging to an arbitrary number of bins. At each time t, a element sm_i corresponds to the information source SM_i and the following notation can be used: $SM_i(t) = sm_i$.

The conditional entropy for two information sources SM_i and SM_j can be calculated as

$$H(SM_j|SM_i) = -\sum_{sm_i}\sum_{sm_j} p(sm_i, sm_j)\log_2 p(sm_j|sm_i) \tag{16}$$

where $p(sm_j|sm_i) = p(sm_j, sm_i)/p(sm_i)$.

$H(SM_j|SM_i)$ is traditionally interpreted as the uncertainty associated with SM_j if the value of SM_i is known.

We can measure synchronicity $s(SM_j, SM_i)$ between two information sources in various manners. One of them is Crutchfield's normalized information distance (which is a metric) between two information sources, defined as (Crutchfield, 1990):

$$d(SM_j, SM_i) = \frac{H(SM_i|SM_j) + H(SM_j|SM_i)}{H(SM_i, SM_j)} \tag{17}$$

Based on this definition we can define synchronicity as

$$s(SM_j, SM_i) = \frac{C}{d(SM_j, SM_i)} \tag{18}$$

We can define the reward associated with a given recent time window as

$$r(SM(\rightarrow t)) = C \cdot (\sum_i \sum_j s(SM_j, SM_i)) \tag{19}$$

Although generally not as a motivational variable, synchrony measures have been used in several recent formal models (e.g. (Prince et al., 2003)). Some related investigations, based on information theory but not specifically to synchrony, have been conducted in which it was studied how various information-theoretic cost functions to be optimized by a sensorimotor system allowed the self-organization of various coordinated behaviour (Sporns and Lungarella, 2006).

4. Conclusion

Based on Berlyne's theory, this paper has proposed an integrated definition of intrinsic motivation, based on the concept of collative variables. Starting from this definition, we have described a landscape of possible computational approaches to intrinsic motivation. Some of them have already been implemented and tested in the literature, and this unified landscape will help to compare them in the same framework. We also showed that there are important types of approaches, such as competence based intrinsic motivation, which are still largely unexplored but full of potential. As a consequence, we hope this paper will set the stage for research initiatives investigating in a systematic manner what are the behavioral and developmental consequences of using each particular type of intrinsic motivation mechanism, in particular studying how far each of them can drive efficiently the learning of reusable skills, self-organize developmental trajectories, and allow for open-ended development.

5. Acknowledgements

We would like to thank Andy Barto for drawing our attention on several key references in the psychology literature on intrinsic motivation.

References

Andry, P., Gaussier, P., Nadel, J., and B., H. (2004). Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms. *Adaptive behavior*, 12(2):117–138.

Barto, A. and Simsek, O. (2005). Intrinsic motivation for reinforcement learning systems. In *Proceedings of the Thirteenth Yale Workshop on Adaptive and Learning Systems.*

Barto, A., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, Salk Institute, San Diego.

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

Berlyne, D. (1965). *Structure and Direction in Thinking.* New York: John Wiley and Sons, Inc.

Bonarini, A., Lazaric, A., and Restelli, M. (2006). Self-development framework for reinforcement learning agents. In *Fifth International Conference on Development and Learning (ICDL).*

Capdepuy, P., Polani, D., and Nehaniv, C. (2007). Maximization of potential information flow as a universal utility for collective behaviour. In *Proceedings of the 2007 IEEE Symposium on Artificial Life*, pages 207–213.

Crutchfield, J. P. (1990). Information and its metric. In Lam, L. and Morris, H. C., (Eds.), *Nonlinear Structures in Physical Systems – Pattern Formation, Chaos, and Waves*, pages 119–130. Springer Verlag.

Csikszentmihalyi, M. (1991). *Flow-the psychology of optimal experience.* Harper Perennial.

De Charms, R. (1968). *Personal causation: the internal affective determinants of behavior.* Academic Press, New York.

Deci, E. and Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior.* Plenum Press.

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

Fedorov, V. (1972). *Theory of Optimal Experiment.* Academic Press, New York, NY.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, Row, Peterson.

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

Huang, X. and Weng, J. (2002). Novelty and reinforcement learning in the value system of developmental robots. In Prince, C., Demiris, Y., Marom, Y., Kozima, H., and Balkenius, C., (Eds.), *Proceedings of the 2nd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems*, pages 47–55. Lund University Cognitive Studies 94.

Hull, C. L. (1943). *Principles of behavior: an introduction to behavior theory.* New-York: Appleton-Century-Croft.

Hunt, J. M. (1965). Intrinsic motivation and its role in psychological development. *Nebraska symposium on motivation*, 13:189–282.

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

Kaplan, F. and Oudeyer, P.-Y. (2003). Motivational principles for visual know-how development. In Prince, C., Berthouze, L., Kozima, H., Bullock, D., Stojanov, G., and Balkenius, C., (Eds.), *Proceedings of the 3rd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems*, pages 73–80. Lund University Cognitive Studies 101.

Marshall, J., Blank, D., and Meeden, L. (2004). An emergent framework for self-motivation in developmental robotics. In *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, Salk Institute, San Diego.

Merrick, K., M. M.-L. (2008). Motivated learning from interesting events: Adaptive, multitask learning agents for complex environments. *Adaptive Behaviour.*

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

Oudeyer, P.-Y. and Kaplan, F. (2006). Discovering communication. *Connection Science*, 18(2):189–206.

Oudeyer, P.-Y. and Kaplan, F. (2007). What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1:6.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(1):265–286.

Prince, C., Hollich, G., Helder, N., Mislivec, E., Reddy, A., Salunke, S., and Memon, N. (2003). Taking synchrony seriously: A perceptual-level model of infant synchrony detection. In Berthouze, L., Kozima, H., Prince, C., Sandini, G., Stojanov, G., Metta, G., and Balkenius, C., (Eds.), *Proceedings of the Fourth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems.* Lund University Cognitive Studies 117.

Roy, N. and McCallum, A. (2001). Towards optimal active learning through sampling estimation of error reduction. In *Proc. 18th Int. Conf. Mach. Learn.*

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

Schembri, M., Mirolli, M., and Baldassare, G. (2007). Evolving internal reinforcers for an intrinsically motivated reinforcement learning robot. In Demiris, Y., Scassellati, B., and Mareschal, D., (Eds.), *Proceedings of the 6th IEEE International Conference on Development and Learning (ICDL2007).*

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

Sporns, O. and Lungarella, M. (2006). Evolving coordinated behavior by maximizing information structure. In Rocha, L. and al., (Eds.), *Proceedings of the 10th International Conference on Artificial Life (Alife'06).*

Sutton, R. and Barto, A. (1998). *Reinforcement learning: an introduction.* MIT Press, Cambridge, MA.

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

Thrun, S. (1995). Exploration in active learning. In Arbib, M., (Ed.), *Handbook of Brain Science and Neural Networks.* MIT Press, Cambridge, MA.

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