

# Interactional Motivation in Artificial Systems: Between Extrinsic and Intrinsic Motivation

Olivier Georgeon, James Marshall, Simon Gay

## ► To cite this version:

Olivier Georgeon, James Marshall, Simon Gay. Interactional Motivation in Artificial Systems: Between Extrinsic and Intrinsic Motivation. International Conference on Development and Learning (ICDL-Epirob), Nov 2012, San Diego, CA, United States. IEEE, pp.1-2, 2012, Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on. <<http://www.icdl-epirob.org/>>. <10.1109/DevLrn.2012.6400833>. <hal-01353128>

HAL Id: hal-01353128

<https://hal.archives-ouvertes.fr/hal-01353128>

Submitted on 13 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Interactional Motivation in Artificial Systems: Between Extrinsic and Intrinsic Motivation

Olivier L. Georgeon, James B. Marshall, and Simon Gay

**Abstract**—This paper introduces **Interactional Motivation (IM)** as a way to implement self-motivation in artificial systems. An interactionally motivated agent selects behaviors for the sake of enacting the behavior itself rather than for the value of the behavior’s outcome. IM contrasts with extrinsic motivation by the fact that it defines the agent’s motivation independently from the environment’s state. Because IM does not refer to the environment’s states, we argue that IM is a form of self-motivation on the same level as intrinsic motivation. IM, however, differs from intrinsic motivation by the fact that IM allows specifying the agent’s inborn value system explicitly. This paper introduces a formal definition of the IM paradigm and compares it to the reinforcement-learning paradigm as traditionally implemented in Partially Observable Markov Decision Processes (POMDPs).

**Index Terms**—Developmental learning, self-motivation, constructivist learning, autonomous agents.

WE propose a new form of self-motivation for artificial agents and robots: interaction-centered motivation, or, more concisely, *interactional motivation* (IM). In essence, we define an IM mechanism as the association of (a) a value function associated with the possible *interactions* that exist between the agent and the environment, and (b) an unsupervised learning mechanism that learns to select behaviors that maximize this value function over time. The key difference with traditional reinforcement learning resides in the fact that the value function is a function of behaviors (interactions) rather than of states.

Motivation is generally defined as a driving force that initiates and directs behaviors. It is widely acknowledged that an agent that selects behaviors to maximize a value function over time conforms to this definition, and, therefore, implements a form of motivation. Typically, a Partially Observable Markov Decision Process (POMDP) can be used to model an agent that learns to perform actions in search for rewarding situations [1]. The designer of the POMDP defines a set of states  $S$ , a set of actions  $U$ , and a reward function  $r : S$

$\rightarrow \mathbb{R}$  that assigns values to states. The POMDP framework is represented in Figure 1.

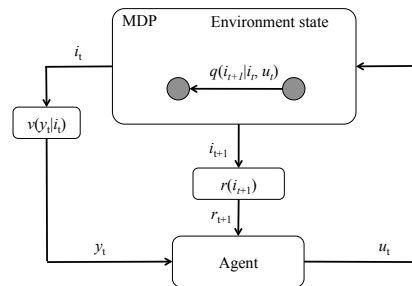


Figure 1: Diagram of a POMDP showing the underlying Markov Decision Process (MDP), and the stochastic process  $v(y_i|i_i)$  mapping the state  $i_i$  (resulting from action  $u_{i-1}$ ) to an observation  $y_i$ , thus partially hiding the environment’s state from the agent.

The designer of the POMDP interprets a state  $i \in S$  as a specific configuration of an agent in a virtual environment, and interprets  $u$  as an action that the agent performs in the environment according to the agent’s policy. In this case, the agent’s motivation is said to be *extrinsic* because the reward function  $r(i)$  is defined independently from the agent’s policy. Moreover, from the designer’s perspective, the agent’s motivation seems to come from something rewarding in the environment that is external to the agent itself.

Extrinsic-motivation models, such as POMDPs, would in fact constitute models of animals that learn to act based on states of the world that have been identified *a priori* by someone else, if such animals existed. Here we argue that animals face a very different problem: the problem of identifying useful states of the world, with regard to the animal’s own motivation. This problem is related to the symbol-grounding problem: grounding the agent’s knowledge in its own activity [5]. Especially in the field of developmental robotics, we expect the meaning of the agent’s knowledge to derive from the agent’s own motivations and experiences rather than from a provided interpretation of the world [10].

The notion of intrinsic motivation has been proposed as a way to implement motivation independently from any specification of the environment. Examples of intrinsic motivation systems are curiosity [7], search for predictability and control [2], and search for simplification of knowledge or compressibility of data [9]. With intrinsic motivation, the value system is generally not made explicit in the form of numerical values. Instead, it is implicitly characterized by the resulting behavior of the system.

This work was supported by the French *Agence Nationale de la Recherche* (ANR) contract ANR-10-PDOC-007-01 and by a research fellowship from the Collegium de Lyon.

O. L. Georgeon is with the LIRIS lab, CNRS, UMR5205, Université de Lyon, France (+33 4 72 04 63 30, olivier.georgeon@liris.cnrs.fr).

J. Marshall is with the Computer Science Department, Sarah Lawrence College, Bronxville, NY 10708 (jmarshall@slc.edu).

S. Gay is with the LIRIS laboratory, CNRS, UMR5205, Université de Lyon, France (simon.gay@liris.cnrs.fr).

In between extrinsic and intrinsic motivation, we propose here the notion of interactional motivation as a way to specify an inborn value system without referring to the environment. Figure 2 illustrates the IM paradigm with a formalism similar to the formalism used in figure 1, to allow for the comparison with POMDPs.

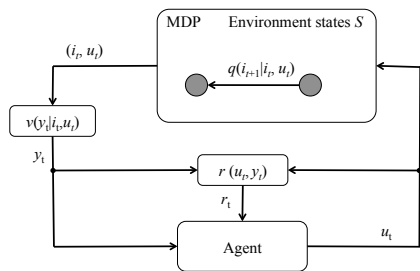


Figure 2: The Interactional Motivation framework. Here, the interaction cycle starts with the agent rather than with the environment.  $y_t$  reflects the feeling resulting from performing action  $u_t$  in state  $i_t$  rather than reflecting the resulting state  $i_{t+1}$ .  $r$  is a function of  $u_t$  and  $y_t$  rather than of  $i_{t+1}$ .

In the IM paradigm, the reward  $r$  is a function of the action  $u$  and of the observation  $y$  rather than of the state  $i$ . We consider the association of an action with the observation that results from that action to be a *sensorimotor pattern* ( $u_t, y_t$ ). Sensorimotor patterns are particularly well exemplified by tactile interactions. Indeed, a tactile interaction is the combination of a movement and the sensory stimulation generated by the movement. We have shown in previous studies [3] that a similar approach can be taken to model vision, following Merleau-Ponty’s idea that vision is “a palpation with the look” [6]. In the following, we refer to a sensorimotor pattern ( $u, y$ ) by the term *scheme* proposed by Piaget in his constructivist epistemology [8].

Conceptually, the agent enacts schemes for their own sake rather than for the value of the outcome that they produce. To make this concept concrete, consider an agent that is motivated to find food. The traditional state-based modeling approach consists of attributing value to states where the agent has reached food. In contrast, the IM approach consists of defining a scheme that consists of eating food, and attributing a positive value to this scheme (a biting action with a good-taste observation). In the former case, the designer defines *a priori* what entities in the world constitute food for the agent, whereas, in the latter case, the agent is left to discover what entities in the world afford the interaction of *eating* on its own.

We define the notion of *enacting a scheme* as performing action  $u$  and obtaining observation  $y$ . Note that an intention to enact scheme ( $u, y$ ) may in fact result in enacting scheme ( $u, y'$ ) if the resulting observation  $y'$  is different from the expected observation  $y$ . The agent may intend to enact ( $u, y$ ) that has a positive value but end up enacting ( $u, y'$ ) that has a negative value. Consequently, the agent must construct knowledge of the world so that it can predict the consequences of enacted schemes and seek situations that lead to positive schemes and avoid situations that lead to negative schemes. This learning process is open-ended because the agent has no goals predefined as reward states.

Another difference with POMDPs is that the distribution  $v$  is now a function of the state  $i_t$  and the action  $u_t$ , rather than of the resulting state  $i_{t+1}$ . With this new definition,  $v(y_t|i_t, u_t)$  specifies the observation  $y$  obtained after performing a given action  $u$  in a given state  $i$ . This difference is important because it means that the observation  $y$  does not always inform the agent about the same property of the environment each time, in contrast to observations in a POMDP. For example, we have implemented IM agents for which the observation is a single bit [4]. At each time step, the meaning of the observational bit  $y_t$  depends on the action  $u_t$  that produced it. When  $u_t$  corresponds to *feel to the left* then  $y_t$  informs the agent about the presence or absence of an object on the left; when  $u_t$  corresponds to *try to move forward* then  $y_t$  informs the agent about the success or failure of this attempt (depending on a possible obstacle). However, the agent initially ignores this meaning. We expect IM agents to demonstrate that they learn this meaning by progressively adapting their behavior to maximize the value function  $r$ . We show examples of such agents online (<http://e-ernest.blogspot.fr/>).

## REFERENCES

- [1] Aström, K. 1965. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications* (10). 174-205.
- [2] Blank, D., Lewis, J. and Marshall, J. 2005. The multiple roles of anticipation in developmental robotics. In *Proceedings of AAAI 2005 Fall Symposium: From Reactive to Anticipatory Cognitive Embodied Systems* (Menlo Park, CA), AAAI Press, 8-14.
- [3] Georgeon, O.L., Cohen, M. and Cordier, A. 2011. A model and simulation of early-stage vision as a developmental sensorimotor process. In *Proceedings of Artificial Intelligence Applications and Innovations* (Corfu, Greece, 2011), 11-16.
- [4] Georgeon, O.L. and Ritter, F.E. 2012. An intrinsically-motivated schema mechanism to model and simulate emergent cognition. *Cognitive Systems Research*, 15-16. 73-92.
- [5] Harnad, S. 1990. The symbol grounding problem. *Physica D*, 42. 335-346.
- [6] Merleau-Ponty, M. 1976. *Phénoménologie de la perception*. Gallimard, Paris.
- [7] Oudeyer, P.-Y., Kaplan, F. and Hafner, V. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11 (2). 265-286.
- [8] Piaget, J. 1970. *L'épistémologie génétique*. PUF, Paris.
- [9] Schmidhuber, J. 2010. Formal theory of creativity, fun, and intrinsic motivation. *IEEE Transactions on Autonomous Mental Development*, 2 (3). 230-247.
- [10] 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 (5504). 599-600.