

Integration of Action, Joint Action and Learning in Robot Cognitive Architectures

Mehdi KHAMASSI*, Benoît GIRARD^(a), Aurélie CLODIC^(b),
Sandra DEVIN^(c), Erwan RENAUDO^(d), Élisabeth PACHERIE^(e),
Rachid ALAMI^(f) & Raja CHATILA^(g)

ABSTRACT. In contrast with research work in Artificial Intelligence, which can test designed reasoning algorithms on specific problems in isolation during perfectly controlled discrete simulations, Robotics always faced the requirement of integrating perception, decision and action execution to work on embodied platforms interacting in and with the real world. This forced roboticists to address in various ways the question of how to efficiently coordinate different memory systems, different levels of decision-making and different learning processes within proposed artificial cognitive architectures. Such robotic proposals have often entertained strong interactions with similar questions as they are also raised in Philosophy, Neuroscience, Psychology and Cognitive Sciences in general. Here we focally review robotic work integrating different levels of action (from goal-oriented action planning to reflexes; from action monitoring to reactive visually-guided behavior) and their associated learning mechanisms. We show that such integration is necessary and sufficient to have an artificial agent fulfill a basic level of self-performance monitoring, and as a consequence to display higher behavioral flexibility, autonomy and generalizability between different environments. We illustrate these issues with examples of experimental work on coordination of multiple learning systems in a single robot and on joint action during human-robot cooperation, finding that partly similar mechanisms for the integration of different action levels can be at work for both individual action and joint action. We highlight some of the successes and failures of these robotic approaches with the hope of feeding and contributing to similar debates on action in other fields of Cognitive Sciences.

Keywords: Action, joint action, learning, planning, human-robot interaction, cognitive architectures, shared plan, shared goals, model-based and model-free reinforcement learning, decision-making.

RESUME. Intégration de l'action, de l'action conjointe et de l'apprentissage dans les architectures cognitives robotiques. À l'opposé des recherches en Intelligence Artificielle, où des algorithmes conçus pour des problèmes spécifiques peuvent être testés dans des conditions de simulation parfaitement contrôlée, la Robotique a toujours eu à faire face à la nécessité d'intégrer la perception, la décision et l'exécution de l'action pour pouvoir fonctionner sur des plateformes physiques en interaction dans et avec le monde réel. Cette particularité a obligé les roboticiennes et roboticiens à se poser très tôt la question de comment coordonner efficacement

* Sorbonne Universités, UPMC, CNRS UMR 7222, ISIR, Paris. mehdi.khamassi<at>isir.upmc.fr.

^(a) Sorbonne Universités, UPMC, CNRS UMR 7222, ISIR, Paris. benoit.girard<at>isir.upmc.fr.

^(b) Univ de Toulouse, CNRS, LAAS, Toulouse. aurelie.clodic<at>laas.fr.

^(c) Univ de Toulouse, CNRS, LAAS, Toulouse. sdevin<at>laas.fr.

^(d) Sorbonne Universités, UPMC, CNRS UMR 7222, ISIR, Paris. benoit.girard<at>isir.upmc.fr.

^(e) Institut Jean Nicod, CNRS UMR 8129, ENS, EHESS, Paris. elisabeth.pacherie<at>ens.fr.

^(f) Univ de Toulouse, CNRS, LAAS, Toulouse. aurelie.clodic<at>laas.fr.

^(g) Sorbonne Universités, UPMC, CNRS UMR 7222, ISIR, Paris. raja.chatila<at>isir.upmc.fr.

différents systèmes de mémoires, différents niveaux de prise de décision et différents processus d'apprentissage dans des architectures cognitives. En conséquence, les propositions d'architectures robotiques ont souvent été confrontées et nourries par les questionnements sur les architectures cognitives tels qu'abordés en Philosophie, en Psychologie, en Neurosciences, et plus généralement en Sciences Cognitives. Dans cet article, nous passons en revue les travaux robotiques qui ont abordé le problème de l'intégration de différents niveaux d'action et de leurs mécanismes d'apprentissage associés (de la planification des actions orientées vers un but aux simples réflexes ; de la surveillance de l'action aux comportements réactifs guidés par la vision). Nous montrons qu'un tel type d'intégration est nécessaire et suffisant pour permettre à un agent artificiel de montrer un niveau basique de surveillance de sa propre performance, et en conséquence de montrer de plus grandes capacités de flexibilité comportementale, d'autonomie et de généralisation à différents environnements. Nous illustrons cette problématique à travers des exemples expérimentaux sur la coordination de systèmes multiples d'apprentissage au sein d'un même robot, et sur l'application à l'action conjointe lors de la coopération homme-robot. Nous trouvons en effet que des mécanismes partiellement similaires d'intégration des niveaux d'action peuvent fonctionner à la fois pour l'action individuelle et pour l'action conjointe. Enfin, nous mettons en lumière certains des succès et des échecs de ces approches robotiques en espérant nourrir et contribuer aux débats similaires sur l'action qui se posent dans les autres champs des Sciences Cognitives.

Mots-clés : Action, action conjointe, apprentissage, planification, interaction homme-robot, architectures cognitives, plan partagé, but partagé, apprentissage par renforcement fondé sur un modèle ou sans modèle, prise de décision.

I – INTRODUCTION

“Im Anfang war die Tat!”¹ Goethe could not put a stronger wording in the mouth of Faust (Goethe, 1808). And it is striking that a neurophysiologist starting in 1993 to head the Chaire de Physiologie de la Perception et de l'Action at Collège de France, Professor Alain Berthoz, also chose this expression² to start his inaugural lesson (Berthoz, 1993). This is a committed position to which roboticists have always been experimentally constrained, in contrast to research in classical Artificial Intelligence where programs were based on perception (*e.g.* a set of information about the world taken as input) and abstract simulated reasoning generating a solution without implying the least action nor the interaction with the environment. *A contrario*, the robot is before anything a body which acts, which physically interacts with its environment.

Robots'ancestors, the automata, critically remind us this feature: without perceiving anything, the purely physical combination of morphological and mechanical properties of the body of Vaucanson's duck, Al-Jazari's peacocks or Hero of Alexandria's Heracles make them move and interact with material objects of the environment (Meyer & Guillot, 2008; Meyer, 2015). More recent robotic researches on “passive dynamic walkers” highlight that without any computer program nor sensor, a couple of metal legs with appropriate “knee-like” joints put on an inclined plane can result in a smooth rhythmic walking behavior which looks to the observer more natural and human-like than most

¹ “In the Beginning was the Act!”

² “Au commencement était l'action.”

existing humanoid robots pre-programmed to walk (McGeer, 1990; Collins *et al.*, 2001). This beautifully illustrates that the body can contribute to action and maybe, as others have suggested, even to cognition (Pfeifer & Pitti, 2012; Vernon, 2014). In line with this, Warnier (2010) wrote that “one’s body ‘is itself one’s actions’ within one’s material environment”.

Nevertheless, the robot cannot be reduced to an automaton (Meyer, 2015). In another inaugural lesson at Collège de France, this time for the first Robotics chair in this institution 50 years after the commonly accepted birth of Robotics on General Motors’ assembly lines in 1961, Jean-Paul Laumond reminds that “Robotics deals with the relation between the real world and a moving machine whose movements are controlled by a computer. As a consequence, the robot shall be distinguished from both the automaton – whose movements are mechanistically determined – and the computer – which manipulates information without moving.”³ (Laumond, 2012).

But the occurrence of movement does not suffice to characterize an action. Roboticists have long ago integrated concepts from philosophical action theory which specifies that an action requires some intention to reach a goal, or at least some kind of mental event playing a causal role in triggering the action. For instance, David MacFarland (1995) proposes the following example to distinguish a movement from an action: “In putting his hands forward to break a fall, John may follow the same pattern of movement as when he puts his hands forward to perform a somersault. The latter is an action, because some mental event (the nature of which depends upon the variety of action theory) played a causal role in the behavior. The former is not an action, because this (special) kind of mental event is lacking (note that there must have been some kind of neutral events that caused the hands to move).” Elisabeth Pacherie (2012) would add that the hand movement may have been caused “by a muscle spasm or by somebody pulling a string attached to [John’s] wrist”. This suggests a wide variety of different possible movements and action types, to which Psychologists have added the distinction between teleological goal-directed behaviors and stimulus-triggered behavioral responses, the latter still not being simple reflexes in that they can be learned through conditioning to seek positive outcomes and avoid negative ones (Dickinson, 1985; Toates, 1995). This distinction maps onto one drawn in Neuroscience between declarative and procedural memory systems, respectively (O’Keefe & Nadel, 1978; Khamassi & Humphries, 2012). Animals thus have a rich repertoire of different types of actions and movements, and their brains are able to coordinate them so that, according to circumstances, the very same example of observed behavior might be a reflection of different learning processes (Adams, 1982) or even to their cooperation (Dollé *et al.*, 2008, 2010a).

³ Nevertheless, the fact that the robot is controlled by a computer program - even a probabilistic one, even a complex one, even one hardly predictable by humans - may suggest to consider the robot’s movements as much “mechanistically” controlled as the automaton. It may also invite to consider the same hypothesis for humans’ movements in a reductionist view of human cognition. Within this view, one would consider the robot and the human as too different higher degrees of complexity on a single continuum starting from the automaton (or from even simpler organisms).

While it can be claimed that initial researches in Artificial Intelligence (AI) have concentrated their efforts on reproducing high-level cognition, reasoning and goal-directed decisions while neglecting the richness and flexibility of low-level sensorimotor learning of actions and movements (Brooks, 1991a) – somewhat trying to build the second story of a house before building the first one (Roitblat & Meyer, 1995) – this paper will argue that Robotics has from its early foundations been constrained to integrate different action levels. For instance, to enable a robot to interrupt a pre-computed action plan in order to avoid a physical obstacle, or to opportunistically orient towards an unexpected proximal resource (Girard *et al.*, 2005). We will review how research in Robotics has progressively converged towards cognitive control architectures for the integration and coordination of perception, decision, action and movement, enabling robots to work in the real-world equipped with sensors and actuators. We will try to convince the reader that the now commonly-used three-layered robot architectures show strong resemblance with distinctions made in Philosophy, Psychology and Neuroscience, suggesting possible exchanges between disciplines. We will illustrate various successes and failures experienced by robotic research on the integration of different action levels, putting an emphasis on the distinction between joint and individual action, with the hope to feed current debates in other fields of Cognitive Sciences. Finally, while robot cognitive architectures so far mostly focus on the coordination of different action levels without incorporating learning mechanisms (because it is already difficult to coordinate high-level action planning with pre-wired low-level reflexes), including the coordination of parallel learning processes in these architectures is the next challenge. We will discuss recent work aiming at integrating learning abilities for different action levels within separate components for declarative and procedural memories of robot cognitive architectures. We will show how the results obtained through Robotics experimentation within this paradigm leads to slightly different behavioral performance and properties than those suggested by related Neurobiological hypotheses, thus offering a way to exploit robotic platforms as experimental tools to refine biological and psychological hypotheses.

2 – INTEGRATION OF PERCEPTION-DECISION-ACTION IN ROBOTICS

One of the first attempts to make autonomous a machine controlled by a computer dates from the late 60s at the Stanford Research Institute (now SRI International). There the Shakey mobile robot was equipped with the STRIPS (Stanford Research Institute Problem Solver) planning algorithm which allowed the robot to reason about the world based on first-order predicate logic. Is object O_1 in room R_1 ? If yes then go to room R_1 , take object O_1 and move it to room R_2 . Although it is an application of symbolic AI to Robotics and is initially based on sequential reasoning under the form of SENSE-MODEL-PLAN-ACT, this real-world application required the integration of perception, planning and action, and went thus beyond AI. The robot had first to perceive the world before being able to reason about it. It was thus equipped with sensors. In addition, the robot had to act and move within the world, and was thus equipped with actuators. More strikingly, the algorithm controlling the Shakey robot combined low-level actions such as moving forward, turning or planning a path, with higher-level actions combining low-level actions. In

addition, the algorithm had to be able to cope with cases of uncertainty or ambiguity on the perception, and could not simply trigger a new action plan if the location of the robot or the identity of the perceived object were not clear enough to ensure a plan can be executed without failure.

In the continuity of this seminal work, numerous planning and high-level deliberation algorithms have been developed and extended for autonomous robots (see Hertzberg & Chatila, 2008; Ingrand & Ghallab, 2015 for reviews). Even if initial work on deliberation, planning and reasoning in Robotics were based on algorithmic solutions from Artificial Intelligence and are still often based on either first-order predicate logic or Bayesian probability theory, the proposed solutions for reasoning in a robot control system differ from reasoning in other knowledge-based systems in a number of aspects. “A knowledge-based robot is an embedded system, working in closed-loop control, typically in a dynamic environment with quite limited resources; it has to interpret itself its sensor input, and should execute at least some of the actions autonomously that it has judged appropriate. Accordingly, knowledge-based robot control has special needs that are not present in, say, an offline expert system, which gets its input by prompting an experienced human in its domain of expertise and writes its output into some document for more experienced humans to deliberate about and possibly act accordingly if they feel it right.” (Hertzberg & Chatila, 2008).

While this sort of approaches led an important part of the Robotics research community to focus on action planning as a general algorithmic solution for the temporal organization of robots’ actions – some examples, strengths and weaknesses of these approaches being discussed later in this article –, it enabled at the same time to progressively refine the distinction between different robot “deliberate functions”, ranging from planning, acting, monitoring, observing, and learning. As Ingrand and Ghallab (2015) wrote: “Deliberation has to closely interact with sensing and actuating. It does not stop at [nor reduce to task] planning and is not reducible to a single function. It critically requires the coherent integration of several deliberation functions. It involves numerous open problems ranging from knowledge representation, acquisition and verification, to scalable reasoning algorithms and architecture issues.” Roboticists came up with methods for the hierarchical organization of action refinements over several levels of state and action spaces, and found through concrete experimentation necessary links between different deliberation functions, for instance monitoring being dependent on prediction in planning and acting, and vice versa. Interestingly, Robotics research on planning methods historically split into distinct solutions, distinct representations and distinct methods for symbolic action planning and geometric motion planning (Minguez *et al.*, 2008). Somehow echoing the nowadays absence of strong integration between Neuroscience communities studying either “decision-making” or “motor control” (Shadmehr & Wise 2005; Delgado *et al.*, 2011; Glimcher & Fehr, 2013; but see Resulaj *et al.*, 2009), roboticists have long considered that these two problems could be decoupled, action planning producing the high-level task-related plan whose steps requiring motions could be handled by specific motion planners. Nevertheless, Robotics research has progressively understood that while this

separation can work in the face of simple problems, higher environmental complexity makes this decomposition no longer possible: the motion constraints having to be taken into account early in the action planning phase (Ingrand & Ghallab, 2015). While an in-depth review of deliberative methods in Robotics is beyond the scope of this article, we will see in the next section that the coordination of deliberative components is one of the ingredients for success in robot cognitive architectures.

In parallel to research work on robot deliberative control, in the 80s a different perspective on the integration between perception-decision-action in Robotics was taken by a biologist, Valentino Braitenberg, in a book entitled “Vehicles” (Braitenberg, 1986). There, a thought experiment more than a concrete robotic experimentation proposes to imagine a series of small wheeled robots (or “vehicles”) made of two wheels, a small number of light sensors and obstacle sensors, and various schemes of interconnections with simple “wires” between these sensors and actuators. Interestingly, the book demonstrates that these vehicles could display a series of behaviors which may look “complex”, “structured”, or even sometimes “intelligent” in the eye of the human observer. In the simplest cases, wiring with an inhibitory connection an obstacle sensor to a motorized wheel on the opposite side of the robot’s body can make it avoid obstacles. Similarly, putting an excitatory connection between a light sensor and a motorized wheel on the opposite side of the body can make the vehicle follow a light in movement and look as if being attracted by light as would some flying insects be. In subsequent examples, combining different types of sensors with different wiring and different timescales of information propagation led to more complex, emerging and temporally organized behaviors of the vehicles.

Following this work, a series of developments have followed such reactive approaches and behavior-based approaches (so-called when a series of modules are connected each specifically dedicated to the control of one type of behavior) to robot behavioral autonomy, resulting in a clash with AI-inspired Robotics community in the early 90s and in the foundation of “new approaches to Robotics” (Brooks, 1991b). This led to important Robotics development on emerging adaptive behavior through the dynamics of sensorimotor loops between the robot and its environment, which had been mostly neglected by previous Robotics research. Some of the work in this direction took inspiration from enaction theory (Varela *et al.*, 1993) and the idea that a cognitive agent is embedded within a closed loop of interaction with the environment. The agent’s perception and action are intricate, they both contribute to the environment, and interactions contribute to preserving the organization of such a closed system. This also relates to the theory of affordances (Gibson, 1977) and the theory of sensorimotor contingencies (O’Regan & Noë, 2001) where the surrounding space is subjectively perceived by the agent as directly linked to its action capacities (*i.e.* by the actions that it *affords*) and its active exploration of its environment. In the field of Robotics, this approach gave birth to the *animat* community (Meyer & Wilson, 1991; Roitblat & Meyer, 2005), with stronger inspiration from simpler but nevertheless highly adaptive biological animals (*e.g.* Meyer *et al.*, 2005). Some of the work within this community was named *Ecological Robotics* – after Gibson’s ecological

approach with affordances –, extending behavior-based approaches with visual processing such as optic flow for robot to navigate adaptively within the environment (Duchon *et al.*, 1998). Similar to previous claims, according to these authors “sufficient information is available in the robot-environment interaction to control the robot’s behavior without further inference or reconstruction [...] internally representing detailed knowledge of the world” through an internal model of it. In other words “it is more desirable to put the animal in its environment than to put the environment into the animal.”

Although not really tackling the issue of high-level planning and reasoning in robots – and thus remaining complementary to research work on robot deliberative functions – this line of research has convincingly demonstrated that some forms of intelligence without representation (Brooks, 1991a) can be achieved in robots. As Hertzberg and Chatila (2008) wrote: “Clearly, the most sophisticated representation and inference system is of no help for a robot, if the reasoning in terms of symbols cannot be grounded in real sensor data and actuator control. On the other hand, representation and inference have much to offer for improving the performance, the software design, and the user interface of robots, for example in field, service, and intervention applications.”

As we will see in the next section, rather than remaining completely isolated and in opposition, these two communities have progressively converged towards the need to integrate deliberation and reactivity in hybrid cognitive architectures, the latter being a means for amalgamating instances of deliberative with reactive control elements in a robot.

3 – STANDARDISATION OF 3-LAYERED ROBOT COGNITIVE ARCHITECTURES

Roboticians have soon had to deal with complex software systems, in large part due to the need to control various sensors and actuators in real time, in the face of important noise and uncertainty. “Robot systems must work to achieve tasks while monitoring for, and reacting to, unexpected situations. Doing all this concurrently and asynchronously adds immensely to system complexity” (Kortenkamp & Simmons, 2008).

A robot-system architecture organizes the components which correspond to the functions the robot will be able to achieve, at several levels of granularity. It describes how these components are linked and interleaved together, how they interact, what information they exchange, and at which temporal rate. In addition to the particular algorithms implemented in the components, the architectural design will hence determine the robot’s capacities to achieve tasks and to react to events.

Researchers in Cognitive Sciences, Neuroscience and Robotics have addressed the issue of the organization and operation of a system (natural or artificial) capable of performing perception, action, deliberation, learning and communication, up to different levels of development (Alami *et al.*, 1998; Anderson *et al.*, 2004; Brooks, 1986; Lehman *et al.*, 1996; Rosenbloom *et al.*, 1993; Vernon, 2014). The term "cognitive architectures" is commonly used in the AI and cognitive sciences communities to designate the organization of systems designed to model the human mind. Two of the most known long-term

projects that claim generality are SOAR (State, Operator And Result) (Rosenbloom *et al.*, 1993; Lehman *et al.*, 1996) and ACT-R (Adaptive Control of Thought-Rational) (Anderson *et al.*, 2004). SOAR aims at modeling human cognition and is based on Newell's seminal work on theories of cognition (Newell, 1990). Operational knowledge in SOAR is represented by production rules. To achieve a goal, the rules' conditions are matched to a "working memory", of which the content is encoded as sets of {attribute-values}. Learning in SOAR is mainly based on a mechanism called "chunking" (other mechanisms such as reinforcement learning are being added). This process is similar to identifying macro-operators, *i.e.* new rules that abstract the succession of rules selected to achieve a goal.

The general concept in ACT-R is a classical rule-based system. Knowledge about facts and events and their relationships is organized in a declarative memory along with a set of production rules and procedures. The data structures in the memory are called "chunks" but the meaning is quite different from the chunks used in SOAR. The rules are selected according to their matching the chunks in the memory, and to their cost and probability of success. Their actions change the declarative memory or trigger actions in the world. The declarative chunks have also an associated "base level" which increases according to their past selection. This results in using chunks that have already been selected, *i.e.*, that were used in more successful activations of the rules. The costs and success rates of the rules is modified according to the outcome of their execution. This provides for an improvement of the global behavior. Furthermore, there is a "compilation" process that produces new rules from analyzing the chunks involved in goal achievement.

There are many similarities between these two main cognitive architectures. They are in agreement on symbolic representations at high levels of abstractions and they both represent operational knowledge by production rules. They both put emphasis on learning mechanisms based on remembering and abstracting previous successful courses of actions. They also both are not very much concerned with real time operation and building representations from sensory data. More precisely, they both say this is important, but at the same time they don't provide a clear approach for achieving it. The question of linking symbolic and subsymbolic representations is actually not really addressed. From a robotic standpoint, this question is central.

The design of *robot control architectures* is a long-standing issue in Robotics. One general paradigm is based the "perceive-plan-act" cycle, which was criticized for its centralization and the lack of a reactive component. Brooks (1986) proposed the "subsumption" architecture, which falls on the other extreme and introduced a different paradigm known as "behavior-based robotics" which has inspired numerous projects. Here, the emphasis was more on the reactive capacity of the system and on its robustness. The organization is that of fixed layers of "sense-react" behaviors with growing complexity, all acting concurrently, the higher ones subsuming the lower ones, which are operational again when the higher levels are not operating. For example, there may be a low-level behavior that simply drives the robot in random direction, this behavior being always active so that the robot always moves and explores its environment. At a second higher-level of the architecture, a behavior could

take sensor input and detect obstacles so as to override the low-level behavior and steer the robot away from any perceived obstacle. This higher-level behavior would also always remain active, but not sending any signal if the robot's surrounding is clear so that the low-level behavior gets control again in this case. Finally, one can conceive of a third-level behavior which always checks the battery level and motivates the robot to interrupt current low-level behaviors in order to search for an energy source or plan a trajectory towards a known one.

One limitation of such behavior-based architectures is that they become rapidly complex and task specific. The priority of particular behaviors over others is fixed a priori by the experimenter, while we will argue in part 5 that the coordination between different types of behavior should result from autonomous learning.

Another limitation is that the systems based on this paradigm do not include a deliberation component and are not taskable. Some authors such as Arkin and colleagues (Likhachev *et al.*, 2002) extended this paradigm by adding a deliberative component, the main role of which is to arbitrate among the behaviors and to coordinate them (which is somewhat contradictory with the rationale of the initial approach). However, behavior-based architectures have difficulties achieving long-range goals or enabling behavioral optimization (Kortenkamp & Simmons, 2008). An illustration of these limits is sketched in the example proposed by (Kortenkamp & Simmons, 2008): "For example, a behavior-based robot that delivered mail in an office building could easily be built by simply wandering around the office building and having behaviors looking for rooms and then overriding the wandering and entering the office. It was much more difficult to use the behavioral style of architecture to design a system that reasoned about the day's mail to visit the offices in an optimal order to minimize delivery time. In essence, robots needed the planning capabilities of the early architectures wedded to the reactivity of the behavior-based architectures. This realization led to the development of layered, or tiered, robot control architectures."

Hybrid architectures are those that include deliberative *and* reactive components together. There is a wide variety of such architectures developed concurrently over the last twenty years. They decompose the robot system into two or three main levels, having different temporal constraints and manipulating different data representations. One of the earliest integrations of reactivity and deliberation comes from the work on "reactive action packages" which constituted one of the first outlines of a three-layer architecture (Firby, 1989). As other examples, we can cite 3T (Bonasso *et al.*, 1997) or Atlantis (Gat, 1991), which are also three-layer architectures. They comprise a set of "skills" in one layer, sequenced by a component in another layer. The third layer is a planning system that reasons on goal achievement including timing constraints. The Aura architecture proposed by Arkin and Balch (1997), initially focused on navigation, integrates a deliberative component (based on an a priori world model) upon a reactive one allowing rapid and effective reaction to changes occurring within the world. Control is based on the notion of schemas inspired from biology, which represent a generic behavior

associated with a perceptual input. This architecture was later extended to include learning and interaction capabilities.

The LAAS architecture (Alami *et al.*, 1998) identifies three main levels: a functional level composed of all the basic built-in robot action and perception capacities; a decisional level which includes the capacities of producing the task plan and supervising its execution, while being at the same time reactive to events, and at the interface between the decisional and the functional levels, an execution control level, which controls and coordinates the execution of the functions distributed in the modules. The CLARAty architecture (Volpe *et al.*, 2001) proposes an organization of software components along two “axes”: an abstraction axis (*i.e.* from low level hardware to high level software processing) and a functional/decisional axis. This results in an architecture that has two layers (a functional layer and a decisional layer). Yet, in each of these layers, one can have components at a very low level of abstraction (*i.e.* close to the hardware) from which one can build up higher-level software components.

In summary of this part, initially strongly influenced by Artificial Intelligence, Robotics defined concepts for autonomous agent actions, such as is STRIPS (Fikes & Nilsson, 1971), based on means-ends analysis, and tackled the problem of robot control architectures, with the objective of building consistent and efficient robot system structures integrating perception, decision and action capacities, and providing for both deliberation and reactivity. Although several different solutions have been proposed, a large part of the community progressively converged to a three-layered architecture (Gat, 1992; Alami *et al.*, 1998; Muscettola *et al.*, 1998; Nesnas *et al.* (2003); Saridis, 1995; Tambe, 1997), which defines:

- A decision level, which includes the capacities for producing a plan to accomplish a task and for supervising its execution, while being at the same time reactive to events from the next level below. The coexistence of these two features, a time-consuming planning process, and a time-bounded reactive supervisory process poses the key problem of their interaction and their integration to balance deliberation and reaction at the decisional level. Basically, the supervisory component uses the planner as a resource when needed and feeds the next level with the sequence of actions to be executed.
- An execution control level, or executive, which controls and coordinates the execution of functions distributed in operational modules (next level) according to the task requirements to achieve the plan. It is at this level that context-based action refinement is performed.
- A functional level which includes all the basic built-in robot action and perception capacities. These functions are encapsulated into controllable communicating modules that enable to implement data processing and motor control loops (image processing, obstacle avoidance, motion control, etc.). In order to make this level as hardware independent as possible, and hence portable from a robot to another, it is interfaced with the sensors and effectors through a logical robot level.

However, these architectures are rather designed to control autonomous robots mainly for planning and executing predefined tasks (see Ingrand &

Ghallab, 2015 for a review). They have little or no learning capacities – which issue will be discussed in part 5 in this paper –, have difficulties to cope with multiple goals, and in general have no abilities for interaction with other robots or with humans (although there have been extensions of LAAS architecture to include this capacity: Clodic *et al.*, 2008). Nevertheless, we will show in the next section that the organization of robot cognitive architectures echoes distinctions made in other disciplines such as Philosophy, Psychology and Neuroscience, opening bridges for fertile exchanges. Then we will illustrate how we recently took profit of such similar action taxonomies between disciplines to feedback Robotics research with Neuroscience knowledge on the coordination of learning at different action levels.

4 – SIMILITUDE WITH PHILO/NEURO/PSYCHO AND CROSS-FERTILITY

4.a - A possible link with the philosophical theories on three levels of deliberation

Interestingly, Pacherie (2012) proposes a dynamic model of intentions that also distinguishes three main stages in the process of action specification:

- distal intentions level (D-intentions) in charge of the dynamics of decision making, temporal flexibility and high level rational guidance and monitoring of action;
- proximal intentions level (P-intentions) that inherits a plan from the previous level and whose role is to anchor this plan in the situation of action, this anchoring has to be performed at two levels: temporal anchoring and situational anchoring;
- motor intentions level (M-intentions), which encodes the fine-grained details of the action (corresponding to what neuroscientists call motor representations), is responsible for the precision and smoothness of action execution, and operate at a finer time scale than either D-intentions or P-intentions.

This suggests an interesting convergence between a philosophical account of the structure and dynamics of human action and a robot control architecture dedicated to action. From this, it appears relevant to consider whether a similar convergence could be established with regard to joint action. This will be the purpose of investigations presented in Part 6 below.

4.b - Distinctions of action types in Psychology

As briefly sketched in the introduction of this article, researches in Psychology have long ago proposed to distinguish different types of actions. While sketching a complete historical perspective on initial work by Skinner, Tolman and others on animal instrumental conditioning and cognitive abilities is beyond the scope of this article (see Dickinson, 1985; Toates, 1995 for reviews), it is nevertheless interesting to note that researchers in the Psychology of Action have progressively converged on the distinction between stimulus-response behavior and outcome-oriented or goal-oriented behavior (Dickinson, 1985). While in the former, the behavioral response is extrinsically

driven by stimuli perceived in the environment, the latter requires some mental representation of action-outcome contingencies enabling the decision to perform actions that are thought to lead to the animal's desired outcome at a given moment.

Psychologists have proposed to operationalize the distinction between these two types of behaviors by the use of outcome devaluation procedures (Adams, 1982; Dickinson, 1985). This consists in devaluing the outcome for which the animal has been previously working, and then looking to which extent the animal is sensitive to this outcome devaluation by measuring how much it continues to display the same behavior even after devaluation. Typically, the animal first learns to press a lever in response of the presentation of a stimulus to get a rewarding outcome (*e.g.* a food pellet). The animal repeats this for a variable duration: experimenters test moderate training versus extensive training. Then in a second phase, the outcome is devalued, for instance by giving a large amount of food to the animal so that it is satiated. Then in a third phase (just after devaluation), the animal is tested again in presence of the stimulus and the lever. After moderate training, animals turn out to quickly stop pressing the lever, which is interpreted as goal-oriented behavior: animals have a representation of the outcome to which pressing the lever leads (*i.e.* food), they do not want this outcome because they are satiated, so they do not perform the action which leads to this outcome. In contrast, after extensive training animals keep pressing the lever as if it has become a habitual and automatic stimulus-triggered behavioral response.

In terms of brain architectures, it is striking that lesions to different parts of the brain (namely the dorsomedial striatum and the dorsolateral striatum respectively) specifically impair these two different types of behaviors (goal-oriented behaviors and habitual behaviors respectively) (Yin and Knowlton, 2006). In the perspective of the three-layered architectures discussed in this paper, one might then naturally ask what would be the third level considered in this psychological framework. This could relate to the ability to perform cognitive control, that is to switch (some would call it "set-shifting") between different behavioral strategies according to internal mental states and depending on task monitoring and self-performance monitoring (Miller and Cohen, 2001). Strikingly, lesions of different subparts of the rat prefrontal cortex (namely the prelimbic and the infralimbic cortex respectively) do not impair specifically goal-oriented or habitual behaviors, but rather impair the ability to switch from habitual to goal-oriented and from goal-oriented to habitual respectively (Killcross & Coutureau, 2003; Coutureau & Killcross, 2003). This suggests that this part of the brain contributes to a higher (third) level of action control which is important for monitoring the ongoing interaction with the environment and activating or inhibiting the lower levels behavioral systems depending on the current requirement of the task. This will be discussed further in the next paragraph on the distinctions between different levels of action in Neuroscience.

4.c - Distinctions in Neuroscience

Importantly for this article, the notion of cognitive architectures is not considered as an archaic terminology or concept in Neuroscience, but is rather

still a hot topic which is for instance the core subject of a recently published special issue in *Neuron* (Dehaene *et al.*, 2015).

In particular, theories of the prefrontal cortex – the anterior part of the frontal lobes of the brain implicated in planning complex cognitive behaviors – have highlighted the hierarchical organization of decision-making, from motor systems to attentional supervisory systems (Koechlin *et al.*, 2003). Such a hierarchical organization relates to a common notion in the Neuroscience literature: the existence of multiple memory systems, with a key distinction between declarative memory and procedural memory. Procedural memory relates to implicit slow learning of reactive behaviors by means of Stimulus-Response associations. In contrast, declarative memory refers to explicit memories which can be consciously recalled such as facts and events (Ullman, 2004) and which are crucial for high-level flexible decision-making. Declarative memory relies on the hippocampus, a key structure whose impairments in humans results in amnesia (Smith & Kosslyn, 2007), which is important to learn configurations of events and cues to build models of the environment (*e.g.* cognitive maps in the spatial domain). In delayed conditioning tasks – where subjects learn a Stimulus-Response association with a certain delay requiring working-memory – the hippocampus is crucial to acquire memory traces and learn the temporal structure of the task (Clark & Squire, 1998; Shors, 2004). Interestingly, recent neurophysiological data show that the communication between the hippocampus and the prefrontal cortex becomes high just after animals have learned the rule of a decision-making task (Benchenane *et al.*, 2010), and this communication is reactivated during subsequent sleep to enable memory consolidation of the recently acquired behaviors (Peyrache *et al.*, 2009). These results tend to show that high-level goal-directed behaviors relying on declarative memory are intimately related to cognitive control and meta-learning processes, in the sense that they involve performance monitoring, task monitoring and related behavioral regulation processes (Khamassi *et al.*, 2011, 2013).

The study of the way the brain decides when to override automatic or habitual reactive behaviors by goal-directed behaviors has already influenced and is itself influenced by algorithmic developments in the machine learning community. First, Peter Dayan and colleagues have shown that the dichotomy between goal-directed behaviors and habitual behaviors in animals parallels the distinction, in machine learning, between model-based reinforcement learning methods (*e.g.* action planning) and model-free methods (*e.g.* Temporal-Difference Learning (Dayan, 2002; Daw *et al.*, 2005))⁴. By trying to reproduce the dynamics of learning observed in animals performing instrumental conditioning tests which enable to distinguish between the two types of behaviors, they showed that the coordination of the two decision-making systems can be well described by a Bayesian uncertainty-based competition: each system evaluates the uncertainty of its own reward estimations; the system with less uncertainty at a given time is the one that takes hand over behavior (Daw *et al.*, 2005). Such uncertainty-based competition may provide

⁴ But note that (Dezfouli & Balleine, 2012) oppose to the proposed identification of habits with model-free reinforcement learning algorithms.

robust techniques for the coordination of learning modules in robots. Moreover, early applications of this principle in simulated navigating agents show that it enables a trade-off between flexible but energy consuming and slow reaction time planning abilities, versus slowly acquired but low-energy consuming and fast reactive Stimulus-Response behaviors (Dollé *et al.*, 2008, 2010a).

Nevertheless, the results obtained so far through the application of machine learning algorithms to Neuroscience remain fragmentary and incomplete for several reasons. First, the laboratory tasks simulated with these models are most of the time very simple, involving only a few different stimuli, states and actions which are given a priori by the experimenter. Second, the debate concerning the possible mechanisms underlying the coordination between different learning systems is still vivid. The method for learning systems coordination proposed by Daw *et al.* (2005) requires costly computations of the uncertainty in the two systems – while the capacity to learn habits may have emerged through evolution to enable computation saving by avoiding to systematically use the goal-directed system (Killcross & Coutureau, 2003). Moreover the complexity of uncertainty computation within the model-based system makes it exponentially explode with the number of states. In the simple task with six states simulated by Daw *et al.* (2005), this is not a problem. But in more realistic situations with a large number of states, this computation becomes problematic. To cope with this issue, the model of Keramati *et al.* (2011) proposes to only compute the less expensive uncertainty of the model-free system, and to avoid using the model-based system when this uncertainty is low. On the one hand this model enables to save computation time and to explain a substantial set of experimental data. On the other hand, this model relies on the simplified assumption that the model-based system is always more reliable and less uncertain than the model-free system. The former is considered to incorporate “perfect information”. As we will see in the next section on robotic tests of these learning models, this is rarely true in more realistic and embodied situations where a robot continuously moves within its environment and has to integrate multiple unstable and noisy perceptions to estimate the configuration of its surroundings (*e.g.* Caluwaerts *et al.*, 2012). Finally, Dezfouli and Balleine (2012) proposed that the model-based system should be coordinated with an action sequence chunking mechanism for habit learning, in replacement of classical model-free reinforcement learning. Their model explains some existing experimental data, but not others.

The discrepancies between results obtained by these competing computational models may be due to (i) the limited dataset on which each model has been tested; (ii) the simplifications and discretizations of the simulated tasks to reproduce, which prevent from studying the more complex dynamics of each learning system in interaction with the real world; (iii) the absence of a common wide dataset used as a benchmark to compare these different models. We have recently proposed to evaluate and compare some of these computational neuroscience-based algorithms for the coordination of learning systems on a common robot platform having to solve a series of different tasks. Such a partially neuro-inspired robotics research work could contribute both to (1) the improvement of robots’ behavioral flexibility and

decisional autonomy by taking inspiration from the way researchers understand how the brain coordinates different learning systems; and to (2) a better understanding of the properties of different computational neuroscience theories by using the robot as a model for biology and as an experimental platform to test and extend the robustness of biological hypotheses.

5 – COORDINATION OF DIFFERENT LEARNING SYSTEMS WITHIN ROBOT COGNITIVE ARCHITECTURES

This part of the article aims at illustrating work in what appears to us as a more recent and challenging goal in Robotics: finding computational principles for the coordination of multiple learning processes in the different layers of a robot cognitive architecture. This part is complementary to parts 2 and 3 which presented the difficulty relying in finding principles for the coordination of different action levels (yet without learning mechanisms) in robot cognitive architectures. Here we will focus on attempts done in the Neurorobotics community which takes inspiration from the development in Computation Neuroscience of principles for learning. Nevertheless, in complementarity to part 4, we will illustrate how robotic experimental tests of such neuroscience-based principles often lead to slightly different dynamics and properties, compared to perfect simulations. This part will thus serve the purpose of opening a dialogue between disciplines on the understanding of principles for the coordination of multiple learning mechanisms.

Despite major progress accomplished in several aspects of Robotics (*i.e.* perception, navigation, localization, motion and action planning, manipulation, human-robot interaction), most of the current results apply to restricted, pre-defined and well-known situations where robots' decisions only apply to quite simple problems (Siciliano & Khatib, 2008). Moreover, robots learning abilities are still very limited, which requires the injection of prior knowledge by the human in the robot's decision-making system. Early applications of reinforcement learning (RL) algorithms to Robotics (*e.g.* Morimoto & Doya, 2001; Smart & Kaelbling, 2002; Alexander & Sporns, 2002; Krichmar & Edelman, 2002; Arleo *et al.*, 2004; Khamassi *et al.*, 2005, 2006) - some of which being neuro-inspired - produced limited progresses, due to applications to relatively simple problems (with a small number of states and actions), to slowness in learning and to systematic instability observed throughout the learning process. More recent applications of RL to Robotics have permitted to deal with more complex and continuous action spaces, enabling to learn efficient sensorimotor primitives (Peters & Schaal, 2008; Sigaud and Peters, 2010; Kober & Peters, 2011; Stulp & Sigaud, 2013). But none of these approaches have attempted to equip robots with an ability to autonomously coordinate different learning systems through self-supervision and to decide which system should have the control over behavior at any given moment, as the mammalian brain does.

Besides, most of robotic decision-making algorithms are based on planning processes which take into account a great number of information, states, locations and actions (*e.g.* Chatila *et al.*, 1992; Alami *et al.*, 2006; Minguez *et al.*, 2008; Kanoun *et al.*, 2011; Ingrand & Ghallab, 2015). Such approach to decision-making could be seen as similar to what neuroscientists call the

model-based system, except that there is most of the time no learning in the system: the internal model is given a priori to the robot and only the planning, decision-making and execution parts have been addressed. Moreover, such an approach raises the issue of having to deal with high-dimensional state spaces, due to the combinatorial explosion in large-scale applications. Another issue which is worthy of note is the long computation time imposed by the planning system, especially if there are systematic replanning of sequences of actions each time the robot is in the same situation and has to decide how to act. In contrast, mammals are able to use routines in familiar environments, controlled by their habit system which is in competition with their planning (model-based) system. The coordination of the planning system and low-level reactive routines is one of the goals of cognitive architectures developed in Robotics (Alami *et al.*, 1998, Volpe *et al.*, 2001).

Such architectures thus appear as a good direction of research for the coordination of decision-making systems in robots and to autonomously decide which system should take over the robot's behavior at each moment (Likhachev *et al.*, 2002). However, these architectures still lack efficient learning abilities and can thus not produce efficient behavioral adaptation in non-stationary environments. In particular, deliberation models are most of the time hand-specified in different representations and do not include learning (Ingrand & Ghallab, 2015).

The interaction of multiple systems to control the behavior of artificial agents has been investigated in a domain at the boundaries of Neuroscience and Robotics, and mainly in navigation tasks. While some studies were using the robotic paradigm to design and test neurocomputational models of animal behavior (Guazzelli *et al.*, 1998; Foster *et al.*, 2000; Chavarriaga *et al.*, 2005; Dollé *et al.*, 2010a; Keramati *et al.*, 2011), thus contributing to neuroscience, some others took inspiration from the goal-directed/habitual behaviors distinction in neuroscience to improve the efficiency of robots (Girard *et al.*, 2005; Caluwaerts *et al.*, 2012; Jauffret *et al.*, 2013; Renaudo *et al.*, 2014, 2015a), thus contributing to Autonomous Robotics. Even if the goal is different in these two general cases, they share many similarities in their methodology:

- An agent interacts with a temporally and spatially continuous environment. The controls of the agent have thus to be fully specified, from the very raw sensory inputs to the low level motor commands that will allow it act, including, of course, the intermediate cognitive components that will integrate information to make decisions.
- The multiple behavioral systems learn in parallel, meaning that every time an action is executed and its consequences measured, all the systems are allowed to learn from this experience, even if they were not or marginally at the origin of this action. This is a key to the generation of synergistic effects of multiple systems cohabitation, even if it can also lead to competition effects.
- All these studies raise similar and specific questions that have been overlooked by too abstract neuroscience models.

A first point of interest is that it is not reasonable to assume that all the world states and all the operators allowing transitions from one state to another are known a priori. These key basic assumptions of classical planning cannot be assumed in many autonomous robot applications. This is true in navigation, for example, if you want to be able to switch your robot on in a place where it has never been before, without providing a map, and let it discover by himself the reachable places and the topology of the environment. It is even more true when the environment is not stationary, *i.e.* when its internal rules of operation can change without the robot to be informed. This can happen when a human closes a door, preventing a mobile robot without manipulator to use a previously cleared pathway. This is why in the multiple learning system Neuroscience literature, the deliberative module is usually based on model-based reinforcement learning systems, which can cope with the incremental discovery of new states and transitions, as well as with their modifications. This choice, initially dictated by operational necessities of Robotics, has deeper consequences on the mechanisms that will coordinate the choices of the multiple systems. Indeed, it cannot be assumed anymore that the deliberative system has a perfect knowledge of the world and of the task at hand (as has been done, for example, in Keramati *et al.*, 2011), and that the only reason for switching from a goal-directed to a habitual behavior would be to save computations in artificial agents or mental load in animals. Such a simplified assumption that the model-based system is always more reliable and less uncertain than the model-free system is occasionally violated in more realistic and embodied situations such as the robotic experiments that we previously performed (*e.g.* Caluwaerts *et al.*, 2012). Caluwaerts *et al.* (2012) found during their robotic tests that depending on noise, uncertainty and characteristics of the environment and of the robot's perceptual equipment, it turns out that most of the time the model-based system is more efficient than the model-free one in some parts of the environment and vice-versa in other parts.

Another limitation of neuroscience-based computational models for the coordination of model-based and model-free reinforcement learning is that they have mostly been tested in simple tasks involving less than 10 states (Daw *et al.*, 2005; Keramati *et al.*, 2011). In this case, the authors can make the assumption that any new tree-search of actions with the highest values within the world-model in order to determine an action plan can restart from scratch. In their case, this is useful to explain to long reaction times experimentally observed when humans or animals employ a deliberative strategy compared to a reactive one. In robotic contexts however, Renaudo *et al.* (2015b) found that the internal world model progressively built by the robot can reach more than 200 states. In this case, re-planning from scratch each time the robot decides to perform a deliberative action makes it too slow to be efficient in a dynamic environment. One solution could be re-plan from scratch in the model-based system but fix a maximal amount of time the robot can take to make a decision. However, in this case only a small subpart of the tree-search process can be completed and the action plans proposed by the robot's model-based system thus become so suboptimal that the performance obtained are most of the time lower than that obtained when the robot use only its model-free system to decide which action to perform. Alternatively, Renaudo *et al.* (2015b) proposed to make the model-based system memorize the estimated action

values resulting from the previous tree search process, and to start from this memory each time a new tree search is triggered. This resulted in an improvement of the performance of the model-based system in their robotic learning task.

Another interesting issue relates to a common neuroscience point of view, derived from the initial experimental observations in rats (Packard & McGaugh, 1996), that the observed strategy shifts in a new task are stereotypical, with goal-directed behaviors being expressed first until habitual systems are sufficiently trained to relay the former. When the incremental nature of the goal-directed system is taken into account in the neurobotic models considered in this section (Girard *et al.*, 2005; Dollé *et al.*, 2010b), the sequence of behavioral strategy use can be different. For example, when starting from scratch, the agent may first rely on basic behavioral strategies, like exploration and approach of perceived reward sources, while building its internal state and transition model, and then use the goal-directed system when enough knowledge has been gathered so that it becomes useful (Girard *et al.*, 2005).

Most multiple system models of navigation are based on the idea that during a single trial (a single attempt to solve the task at hand), the agent may base each of its decisions about the direction of movement to choose next on a different strategy or set of strategies. Again, it is at odds with neuroscience experiments, where (mainly because of experimental limitations) it is considered that a single behavioral type is expressed during a whole trial (see for example Rondi-Reig *et al.*, 2006). It means that these models all consider, one way or another, that in addition to the multiple behavioral systems, a mechanism permanently evaluates and controls the relative contribution of each system to the expressed behavior. Two main schemes have been proposed for this mechanism: either a fusion of the behavior suggested by each system (Guazzelli *et al.*, 1998; Girard *et al.*, 2005; Renaudo *et al.*, 2015a), which implies that all the behavioral systems are always active and continuously provide suggestions for action, or a selection of the behavioral system to activate (Foster *et al.*, 2000; Chavarriaga *et al.*, 2005; Dollé *et al.*, 2010a,b; Keramati *et al.*, 2011; Caluwaerts *et al.*, 2012a,b; Jauffret *et al.*, 2013; Renaudo *et al.*, 2015a), in which case a selection criterion has to be defined.

The fusion schemes combine the vectors of values associated to each action, provided by each behavioral system, with either sums, weighted sums, multiplications, as well as rank or majority voting methods. The study from (Renaudo *et al.*, 2015a) showed that, in a specific benchmark task, various combination methods (namely multiplication, rank and majority voting) can achieve better performance than a baseline defined as a uniform random choice among systems at each time step. This result is however not general, and such comparison studies should be carried out in other contexts to possibly identify the most successful fusion schemes.

The selection at each moment of one system only to control the behavior has an advantage over fusion: if this choice is done a priori, it means that the computations of the selected system only will be made, hence saving computational power. Of course, a drawback is that information that could have been useful (the suggestions from the other systems) is thrown away. A

simple behavioral measure of lack of progress (frustration) has been proposed (Hasson, 2011; Jauffret *et al.*, 2013), it has the elegance of parsimony, but can only be used when such a progress can be measured by the agent on the task at hand, meaning that it has some a priori knowledge about what it is to complete the task. Most of the other selection criteria have been based on reward information: the system which proved in the past to be the best to provide reward, is favored (Foster *et al.*, 2000; Chavarriaga *et al.*, 2005; Daw *et al.*, 2005; Dollé *et al.*, 2010a,b; Keramati *et al.*, 2011; Caluwaerts *et al.*, 2012; Collins & Frank, 2012). Estimation of the uncertainty of each system's evaluations has also been considered either explicitly (Daw *et al.*, 2005), or implicitly, for example by comparing reward prediction errors of each system (Chavarriaga *et al.*, 2005), by comparing the entropy of their action distributions (Renaudo *et al.*, 2015a), or by making a tradeoff between the value of computing more certain information and the cost of obtaining it (Keramati *et al.*, 2011).

Finally, a hybrid combination scheme has recently been proposed (Pezzulo *et al.*, 2013; Viejo *et al.*, 2015), where the values from the habitual system are combined with those of the goal-directed one (fusion), but where the computation of this goal-directed system are monitored step-by-step, and stopped when the uncertainty has been reduced enough, which can in some cases result in a partial or total inactivation of the goal-directed system (selection).

In these last three neuroscience-oriented papers (Keramati *et al.*, 2011; Pezzulo *et al.*, 2013; Viejo *et al.*, 2015) the cost of activating the goal-directed system is taken into account. However, no robotics-oriented study has, until now, addressed the question of the computational cost of activating a given system, despite the fact that in robots this cost can be explicitly quantified.

To conclude, Neurorobotics and Computational Neuroscience have shown that the incremental acquisition of knowledge may play a role in the choice of a given behavioral system, depending on the current level of task exploration, that strategy selection can be revised at a very small time-scale, and that criteria such as reward rate, uncertainty and computational cost should be taken into account to combine the different systems, even if no definitive algorithm has yet emerged.

All the robotic work presented so far on the integration of learning at different action levels within robot cognitive architectures remains at the level of individual behavior. One important question remains whether similar distinctions between action levels, and similar mechanisms to coordinate/learn these actions are at stake in the case of human-robot interaction, in particular during joint action.

6 – INTERACTION AND JOINT ACTION

For more than a decade, the field of human-robot interaction has generated many valuable contributions of interest to the Robotics community at large. We will here give some insights concerning a particular type of interaction which is joint action. First, we will explain which processes are involved in human-human joint action and then in human-robot joint action.

Successful joint action depends on the efficient coordination of participant agents' goals, intentions, plans, and actions. In other words, it is not enough that agents have a common goal and then that each sets their own sub-goals, devise their own individual action plan and execute this plan. They must also coordinate their own sub-plans with those of their co-agents so as to have a coherent joint action plan and they must coordinate their actions in the execution phase to insure the successful completion of the joint action. For that they must monitor both their own and their partner's intentions and actions, understand their combined impact on their joint goal and adjust what they do accordingly. In other words, agents should be able to align their representations of what they themselves and their partners are doing, and of how these actions together contribute to the shared goal.

There has been a great deal of work in recent years, both conceptually and empirically, investigating the cognitive processes by which uncertainty is reduced and coordination achieved. Philosophical accounts of joint action have tended to concentrate on the processes involved in the formation of shared intentions and in high-level action planning prior to acting. In contrast, cognitive psychology studies of joint action have explored the perceptual, cognitive, and motor processes that enable individuals to flexibly coordinate their actions with others online. Because philosophers and psychologists focus on processes operating at different levels of action specification, it is important to bring together their complementary perspectives to shed light on the whole range of processes involved in acting together.

Philosophical accounts of shared intentions are attempts to cash out what it takes for agents to act in a jointly intentional manner. These accounts typically agree that shared intentions are more than mere summations of individual intentions, but tend to disagree on what more is needed. For instance, according to Bratman's influential account (Bratman, 2014), shared intentions are characterized by a form of mutual responsiveness of each to each in their relevant intentions and plans. In contrast, Margaret Gilbert (2009) takes joint commitments to constitute the core of shared intentions: agents share an intention to do A if and only if they are jointly committed to intend as a body to do A. Finally, Raimo Tuomela (2013) points out that when agents act jointly as members of a group, what he calls *we-mode* joint action, they are often committed not just to a particular joint goal but also to a set of values, standards, beliefs, practices, social coordination conventions, pre-established scripts and routines and so on, that form the ethos of the group. While these philosophers have divergent views regarding the nature of the social glue that binds together the intentions of individuals in joint action, their accounts tend to be cognitively demanding: the coordination processes involved in forming and maintaining a shared intention rest on advanced representational, conceptual and communicational skills and sophisticated forms of reasoning about the complex interplay between each other's individual beliefs and intentions and the shared goal, about the mutual obligations and entitlements the shared intention generates, or about its relations to the group ethos.

In contrast to philosophical approaches, cognitive psychology studies of joint action have tended to focus not on the formation of shared intentions but rather on the perceptual, cognitive, and motor processes that enable individuals

to flexibly coordinate their actions with others online. Following Knoblich and colleagues (Knoblich *et al.*, 2011), we can distinguish between two broad categories of online coordination processes: emergent and intentional. In intentional coordination, agents plan their own motor actions in relation to the joint goal and also to some extent to their partners' actions. Thus, shared task representations, that specify in advance what the respective tasks of each of the co-agents are, provide control structures that allow agents to monitor and predict what their partners are doing, thus enabling interpersonal coordination in real time. For instance, people tend to predict the sensory consequences not only of their own but also of other participants' actions (Wilson & Knoblich, 2005) and to automatically monitor their own and others' errors (van Schie *et al.*, 2004). An important complement to the co-representation of tasks and actions is joint attention, that provides a basic mechanism for sharing perceptual representations of objects and events and thus for creating a perceptual common ground in joint action (Tomasello & Carpenter, 2007). Joint attention can also allow agents to perform joint actions more efficiently. For instance, co-agents engage in a joint visual search task are able to distribute a common space between them by directing their attention depending on where the other is looking and their joint search performance is more efficient than their performance in an individual version of the task (Brennan *et al.*, 2007). Coordination smoothers, *i.e.* any kind of modulation of one's movements that 'reliably has the effect of simplifying coordination' is another type of process that may contribute to better online coordination (Vesper, 2014). For example, one may exaggerate one's movements or reduce variability of one's movements to make them easier for the other participant to interpret.

In emergent coordination, coordinated behavior occurs due to perception-action couplings that make multiple individuals act in similar ways. One source of emergent coordination involves interpersonal entrainment mechanisms. Thus, two people sitting next to each other in rocking chairs will unconsciously synchronize their rocking frequency (Richardson *et al.*, 2007) and two people walking side by side will tend to fall in synchrony (van Ulzen *et al.*, 2008). Another source of emergent coordination is perception-action matching, whereby observed actions are matched onto the observer's own action repertoire and can induce the same action tendencies in different agents who observe one another's actions (Jeannerod, 1999; Prinz, 1997; Rizzolatti & Sinigaglia, 2010). It is likely that such processes make partners in a joint action more similar and thus more easily predictable, and thereby facilitate coordination during action execution.

Humans thus have at their disposal a vast array of processes, ranging from advanced representational, conceptual and communicational skills and sophisticated forms of reasoning to intentional and automatic online alignment processes, that they can use to promote interpersonal coordination. To enable efficient joint action, these processes must work together, as there are complementary limits on what each can do.

Human-robot joint action faces similar coordination challenges. We will explain now a way they can be translated to this case and quote some related implementation.

The robot needs to be able to handle a representation of itself and a representation of the human it interacts with. Doing so, it must be able to infer how each of these representations evolves along the joint action unfolding. The robot has to be able to consider perspective taking ability, knowing that representations evolve differently given each one point of view. Among others, Milliez *et al.* (2014) and Hiatt and Trafton (2010) endow a robot with the ability to construct a representation of other agents mental states concerning the environment allowing it to pass the Sally and Anne test (Wimmer & Perner, 1985). Then, these mental states are used in Hiatt, Harisson and Trafton (2011) to interpret and explain humans behaviour.

But the robot also needs to understand and take into account the effects of its own actions into the mental states of its partners. This is done in Gray and Breazeal (2014) where the robot plays a competitive game with a human and chooses its action in order to manipulate the mental state of the human.

Each agent must have access to situation assessment abilities that will enable it to anchor the situation of action: the objects to be acted upon, their location as well as the location of possible obstacles. In Sisbot, Ros and Alami (2011) the robot uses the geometric information about the humans and the objects to construct symbolic knowledge as humans capabilities (an object is visible or reachable by someone) or relations between objects (an object is on/in another one). Lemaignan *et al.* (2012) use this knowledge to anchor situated discourse during human-robot interaction. For example, if a human point at a mug saying "Give me that mug" the robot is able to understand that the human wants this mug and not another one.

During human-robot joint action, joint attention is also key because it states that if joint attention is established, whatever information I can get, I can consider my interactor would have it too if it occurs in the joint attention space. Staudte and Crocker (2011) show that people interpret robot gaze in a same way as human gaze and that a congruent use of the robot gaze helps its human partner to understand spoken references. Mutle, Terrell and Huang (2013) also show that the use of congruent robot gaze cues with speech references enable more effective joint attention by disambiguating spatial references in speech, improving task performance in joint action and the perceptions of the robot. They also show that robots might improve task performance and user experience in human-robot collaboration by using action observation.

Another capacity needed by the robot, emphasized, among others, by Tomasello *et al.* (2005) as a prerequisite to joint action, is to be able to read its partner's actions. Gray *et al.* (2005) use the concept of mirror neurons, introduced by Gallese and Goldman (1998), to infer human action goals by matching the human movement to one of its own movements (even if the robot's morphology differs from that of the human). Hawkins, Bansal and Bobick (2014) endow a robot with the capability to probabilistically infer what the human is doing and what he will do next in order to anticipate and prepare collaborative action.

Complementary, shared task representations are important. It means, if we paraphrase Knoblich, Butterfill and Sebanz (2011)'s definition, that we must equip the robot with a model of what the respective tasks of each of the co-agents are and also control structures that will allow itself to monitor and

predict what its partners are doing. Nikolaidis and Shah (2012) present a method allowing the robot to build a shared mental model of how to perform a collaborative task by looking at human performing the task and then use it when performing the task with a human.

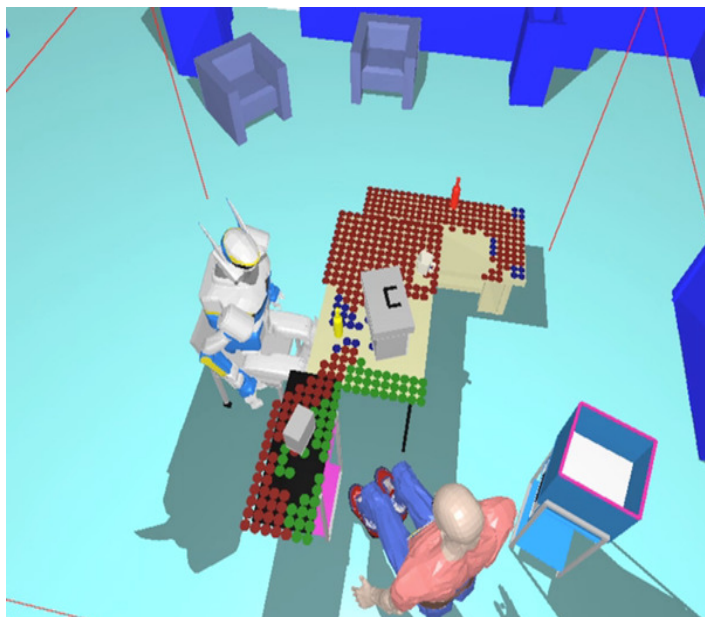


Figure 1

Human-Robot Interaction task performed at the LAAS with the HRP2 robot.

We've seen that coordination constraints apply both to human-human joint action and to human-robot joint action and they can undermine both of them. However, they do not apply with the same strength. In human-human joint action, a human faces another human. The fact that they are both humans brings lots of background knowledge and assumptions can be made from both sides on what the other knows or not. This is far to be easy to assess in the human-robot case.

Nevertheless, we have seen that human-human joint action sometimes involves planned joint action with explicit shared goals, action plans and attentions, and sometimes involve automatic synchronization or alignment processes between partners at a more sensorimotor level. Thus one might reasonably postulate that the integration of learning at different action levels within robot cognitive architectures which has previously been applied to individual robotic tasks – such navigation tasks in Caluwaerts *et al.* (2012) or sequential decision-making tasks in Renaudo *et al.* (2014) – may be relevant in the context of human-robot interaction. This could enable the robot to automatically switch between automatic/habitual behavior and planned action depending on the requirement of the task, and thus display more behavioral flexibility and efficiency during joint action with humans. Renaudo *et al.* (2015c) recently put forward this hypothesis and applied their model-based/

model-free reinforcement learning robot architecture to a human-robot interaction task proposed by Alami *et al.* (2013) and schematically illustrated in Figure 1. In this type of tasks, human and robot have to cooperate to clean a table by putting all objects on the table in a box. The robot has to elaborate a representation of different sub-spaces on the table so that it understands that some objects are accessible to the human (green area in Figure 1), some objects are visible by the human but not reachable by him (red area in Figure 1), while some objects are accessible by the robot. When the robot is endowed with role A (the box being close to it), it can take objects and put them in a box close to him, then ask the human to give him other objects and put them in the box. When the robot is endowed with role B (the box being only accessible to the human), the robot has to put objects in areas accessible to the human so that the latter puts them in the box. Previous work in this type of human-robot task involved pre-defined world models so that the robot can plan complex action plans involving alternations between actions performed by the robot and actions performed by the human with a human-aware task planner (HATP) (Alami *et al.*, 2013). This however did not involve a learning process. Renaudo *et al.* (2015c) proposed an extension of this work by considering that the subparts of the action sequence that are repeatedly performed in the same manner in this condition can be learned by the model-free habit learning system of their architecture. This is similar to habits learned by humans in conditions where repetitive behaviors are always occurring in the same context and in the same manner. This could enable the robot to autonomously detect which situations are stable enough and repetitive enough to avoid systematically using the slow and costly action planning system. In addition, this framework should also enable the robot to automatically detect when environmental changes require to break the habit and switch back to the planning of new action sequences.

As a first step, Renaudo *et al.* (2015c) showed that the combination of model-based planning with HATP and model-free learning within their architecture could increase the learning speed of the robot in comparison with the use of each of these action system in isolation. In the future, they plan to also use model-based reinforcement learning in their architecture to refine the world-model used by HATP in order to find other action plans that could not be anticipated by the human experimenter.

7 – DISCUSSION

In this paper, we have reviewed robotic work aiming at integrating different levels of action control within robot cognitive architectures. We have first taken a historical perspective, describing some early robotic applications of Artificial Intelligence (AI)-based reasoning algorithms to Robotics, other robotic work on the emergence of complex adaptive behaviors through the bottom-up build-up of simple sensorimotor schemes, and on the more recent attempts to combine these two approaches in hybrid architectures. We have showed how the robotic community has progressively converged to a standard on using three-layered architectures, and how this echoes to different distinctions of action types made in Philosophy, Psychology and Neuroscience. This similarity has already resulted in multiple exchanges between disciplines, and suggests additional promising future exchanges and cross-fertility,

including experimental paradigms employing robots as platforms to test hypotheses coming from other disciplines and assessing their robustness on embodied artefacts interacting with the real-world and with humans.

We have tried to illustrate how the embodied and situated aspects of Robotics made robotic work on action go beyond traditional AI in the understanding of the contribution of the body and the role of the integration of multiple action levels (corresponding to different memory systems) in producing flexible behavior. We have in particular illustrated the application to robot architectures of the neuro-inspired coordination of the model-based / model-free reinforcement learning framework, enabling robots to autonomously alternate between learning a world-model and using it to deliberately plan sequences of actions, and avoiding costly action planning when the environment is sufficiently stable to enable behavioral habit learning. We have also illustrated the case of joint action during human-robot interaction, showing that partly similar mechanisms and distinctions between action levels could be at stake during individual and joint action. Finally, we have presented recent robotic work starting from this last consideration to apply architectures integrating learning at different action levels to tasks involving human-robot cooperation. Although this work is preliminary, the results obtained so far suggest that within this paradigm also the neuro-inspired combination of planning and model-free reinforcement learning could speed up the robot's learning capabilities.

We have also tried to illustrate several times throughout this article one of the contributions that Robotics can bring to Neuroscience by using robots as experimental platforms to test biological hypotheses. We have in particular highlighted some results obtained with robotic models for the coordination of model-based and model-free reinforcement learning that could raise new issues and associated solutions not previously raised by computational neuroscience models. Another example which could somehow summarize the approach using robots to test neuroscience-based learning hypotheses is the work of Khamassi *et al.* (2011) which have proposed a robot model of learning with dopamine signalling prediction errors based on salient phasic events. They have started by applying classical reinforcement learning models to reproduce monkey behavioral performance with a robot. However, in contrast to previous simulations of computational neuroscience learning models where the human experimenter can tell the model when is the end of each trial and thus when the model should expect a feedback signal, here the authors decided to avoid giving this information to the robot. The interesting solution proposed consisted in making simulated dopamine neurons phasically respond to any salient event, but only reinforce artificial signals when the dopamine signal is congruent with a motor efference copy - which indicate that an action has been performed and thus that an outcome from this action is expected. This robotic solution is both (i) more parsimonious than previous solutions by avoiding the human experimenter to pre-define when should be the feedback signal; and (ii) closer to some dopamine recordings which contradicted the classical reinforcement theory by showing responses to novel stimuli or even salient aversive signals. A wide array of studies have similarly used robots as platforms to test neuronal architectures composed of multiple brain areas,

studying the dynamical properties of such complex systems during physical interaction with the environment. We invite the reader interested in a more extensive panorama of the variety of contributions of Robotics research to Neuroscience questions to read the recent review by Floreano *et al.* (2014).

Another key aspects of Robotics research on action which might lead to major progress in the forthcoming years is the ability to propose concrete, operational, algorithmic solutions for the online, progressive and autonomous buildup of the robot's own representations relevant for action. Although this is still lacking in most current cognitive architectures and is still a recent topic, it is one of the goals of the field of Developmental Robotics to have a robot explore its environment and discover affordances, relevant descriptions of the perceptual features, states of the environment, actions, goals and subgoals that can be appropriate to satisfy in the long-term both its intrinsic motivations such as the need to minimize uncertainty about the world while maximizing its energy, and extrinsic motivations such as positive feedback from the environment or from humans around (Asada *et al.*, 2001; Baldassarre and Mirolli, 2013; Lungarella *et al.*, 2003; Oudeyer *et al.*, 2007; Schmidhuber, 2006; Weng, 2004).

A particularly illustrative example of recent successful attempt in this direction is the integration of goal representation learning combined with low-level object reaching skill learning in a multi-layered robot cognitive architecture is the work of Santucci *et al.* (2016). In their task, a robot with two arms learns through motor babbling that some movements (*i.e.* touching small spheres in the environment) produces a change in the environment (*i.e.* the spheres light up). If a change occurs consistently several times, the robot can build a goal representation of it in order to later decide to pursue this goal so as to refine the arm movement that can produce it. In addition, the proposed architecture autonomously learns to allocate the appropriate computational resources (neural-networks submodules) and morphological resources (which arm to use for a given goal). Interestingly, although the proposed architecture does not reach optimal performance because it sometimes makes mistakes, not allocating the appropriate resources or the appropriate arm to reach a particular goal, it most of the times reaches satisfying performance and can even sometimes do better than a pre-specified architecture: Strikingly, in a scenario where a sphere located close to the robot's left arm turns out to be more easily reachable by the right arm, this architecture can autonomously learn it while an architecture where the human experimenter has pre-allocated the robot's left arm to reach this sphere would fail. In other words, a solution that could have appeared obvious to a human experimenter designing a dematerialized simulation (*i.e.* considering that objects on the left should be reached by the left arm) may not be the optimal solution in some situations and turns out to be something that needs to be autonomously learned by an embodied agent, pushing the roboticist to propose a computational mechanism/solution for this issue.

This and many similar experimental examples pushed the Robotics community to go beyond early developed systems which assumed that any planned action sequence would be directly executable by the robot, and which had thus reduced the action control process to reacting without much

refinement deliberation. Several different methods have been proposed to overcome this issue, each with advantages and disadvantages, but which presentation is beyond the scope of this paper. Nevertheless, it is interesting to note following Ingrand and Ghallab (2015) that: “Their drawback is the demanding engineering effort required for the specification of skills. They raise significant obstacles for verification and validation, which are mandatory in safety critical applications.” Developmental solutions and any work on the integration of learning capabilities in robot cognitive architectures appear as promising solutions, keeping in mind that the design of appropriate knowledge representations for planning and action execution should integrate as early as possible the requirements of performance monitoring, so that the robot can efficiently coordinate subsystems based on auto-evaluation. Moreover, rather than thinking that only low-level sensorimotor learning solutions should be sufficient for any robot learning task, hybrid architectures tested in developmental paradigms such as learning through goal babbling found that a high-level goal representation, as long as it is autonomously built by the robot, can bootstrap learning based on the advantages of searching solutions within the task space rather than within the larger joint space (*e.g.* Rolf *et al.*, 2011).

A last example of interesting thoughts raised after extensive robotic tests of cognitive architectures concerns the issue of not only coordinating the competition between different action systems within architectures but also making them cooperating. With this respect Mataric and Michaud (2008) note that “as long as the outputs of the two [deliberative and reactive] components are not in conflict, the system requires no further coordination. However, the two parts of the system must interact if they are to benefit from each other. Consequently, the reactive system must override the deliberative one if the world presents some unexpected and immediate challenge. Analogously, the deliberative component must inform the reactive one in order to guide the robot toward more efficient and optimal trajectories and goals. The interaction of the two parts of the system requires an intermediate component, which reconciles the different representations used by the other two and any conflicts between their outputs. The construction of this intermediate component is typically the greatest challenge of hybrid system design.” Interestingly, Engineering research not only suggested to incorporate intermediate components for the construction of shared representations between decisional modules but also to reduce the connection cost by having a single conflict solving component connected to each module rather than a fully interconnected architecture. This principle has already inspired some biologists when proposing that a part of the brain such as the basal ganglia may play the role of a central action selection device in charge of preventing the simultaneous execution of incompatible conflicting actions (Redgrave *et al.*, 1999). Further investigations collaborations between Robotics and Neuroscience will be required to further understand which architectural organization of action control can produce behavioral flexibility and autonomy without the cost of reducing efficiency nor the analyzability of system.

One of the important elements still lacking in the approaches presented here are emotions. From the point of view of Cognitive Sciences, one might think that emotions are key to convey a real meaning to behaviors, and thus to

transform simple behaviors into real actions (Steiner & Stewart, 2010; Di Paolo, 2010). The hybridization of robot cognitive control architecture may benefit in the future from the integration of emotion in decisional processes.

REFERENCES

- Adams, C.D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology*, 34B, 77-98.
- Alami, R., Chatila, R., Fleury, S., Ghallab, M. & Ingrand, F. (1998). An architecture for autonomy. *International Journal of Robotics Research*, 17(4), 315-337.
- Alami, R., Clodic, A., Montreuil, V., Sisbot, E. A., & Chatila, R. (2006). Toward Human-Aware Robot Task Planning. In *AAAI Spring Symposium: To Boldly Go Where No Human-Robot Team Has Gone Before* (pp. 39-46).
- Alami, R., Warnier, M., Guitton, J., Lemaignan, S. & Sisbot, E.A. (2013). When the robot considers the human... In O. Khatib & H. Christensen (eds.) *Proceedings of the 15th International Symposium on Robotics Research*. Berlin-Heidelberg: Springer.
- Alexander, W. H., & Sporns, O. (2002). An embodied model of learning, plasticity, and reward. *Adaptive Behavior*, 10(3-4), 143-159.
- Anderson, J.R., Bothell, D., Byrne, M.D., Douglas, S., Lebiere, C. & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4), 1036-1060.
- Arleo, A., Smeraldi, F., & Gerstner, W. (2004). Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. *Neural Networks, IEEE Transactions on*, 15(3), 639-652.
- Arkin, R. & Balch, T. (1997). Aura: Principles and practice in review. *Journal of Experimental and Theoretical AI*, 9(2).
- Asada, M., MacDorman, K.F., Ishiguro, H. & Kuniyoshi, Y. (2001). Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37(2), 185-193.
- Baldassare, G. & Mirolli, M. (2013). *Intrinsically motivated learning systems: an overview* (pp 1-14). Berlin-Heidelberg: Springer.
- Bangerter, A. & Clark, H.H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, 27, 195-225.
- Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P. & Wiener, S.I. (2010). Coherent theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron*, 66(6), 921-936.
- Berthoz, A. (1993). *Leçon inaugurale*. Collège de France, Paris.
- Bonasso, R.P., Kortenkamp, D., Miller, D. & Slack, M. (1997). Experiences with an architecture for intelligent reactive agents. *Journal of Experimental and Theoretical AI*, 9(2).
- Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. Cambridge, Mass.: The MIT University Press.
- Bratman, M. (2014). *Shared Agency*. Oxford University Press.
- Breazeal, C. (2003). Towards sociable robots. *Robotics and Autonomous Systems*, 42(3-4), 167-175.
- Brennan, S. E., Chen, X., Dickinson, C., Neider, M. & Zelinsky, G. (2007). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106, 1465-1477.
- Brooks, R.A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2(1), 14-23.
- Brooks, R.A. (1991a). Intelligence without representation. *Artificial Intelligence*, 47, 139-159.
- Brooks, R. (1991b). New approaches to robotics. *Science*, 253(5025), 1227-1232.

- Caluwaerts, K., Staffa, M., N'Guyen, S., Grand, C., Dollé, L., Favre-Félix, A., Girard, B. & Khamassi, M. (2012). A biologically inspired meta-control navigation system for the psikharpax rat robot. *Bioinspiration & Biomimetics*, 7(2), 025009.
- Chatila, R., Alami, R., Degallaix, B. & Laruelle, H. (1992). Integrated planning and execution control of autonomous robot actions. In *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA '92* (pp. 2689-2696).
- Chavarriaga, R., Strösslin, T., Sheynikhovich, D., & Gerstner, W. (2005). A computational model of parallel navigation systems in rodents. *Neuroinformatics*, 3(3), 223-241.
- Clark, H.H. (1996). *Using Language*. Cambridge University Press.
- Clark RE & Squire LR. (1998). Classical conditioning and brain systems: the role of awareness. *Science*, 280(5360), 77-81.
- Clodic, A., Cao, Q., Alili, S., Montreuil, V., Alami, R. & Chatila, R. (2008). Shary: a supervision system adapted to human-robot interaction. In *ISER'08 -11th International Symposium on Experimental Robotics*, Athens, Greece.
- Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7), 1024-1035.
- Collins, S., Wisse, M. & Ruina, A. (2001), 'A three-dimensional passive dynamic walking robot with two legs and knees', *International Journal of Robotics Research*, 20, 607-615.
- Coutureau, E. & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioral Brain Research*, 146(1-2), 167-174.
- Daw N.D., Niv Y. & Dayan P. (2005). Uncertainty based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8, 1704-1711.
- Dayan., P. (2002). Motivated reinforcement learning. In TG. Dietterich, S. Becker, & Z. Ghahramani (eds), *Advances in Neural Information Processing Systems 14* (Vol 1 & 2, pp. 11-18). Cambridge, Mass.: The MIT Press.
- Dehaene, S., Dudai, Y. & Konen, C. (2015). Cognitive Architectures. *Neuron*, 88(1), 1.
- Delgado, M.R., Phelps, E.A. & Robbins, T.W. (2011). *Decision making, affect, and learning: Attention and performance XXIII* (Vol. 23). Oxford University Press.
- Di Paolo, E.A. (2010). Robotics inspired in the organism. *Intellectica*, 53-54, 129-162.
- Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society, London, B* 308, 67-78.
- Dollé L., Sheynikhovich D., Girard B., Chavarriaga R & Guillot A. (2010a). Path planning versus cue responding: a bioinspired model of switching between navigation strategies. *Biological Cybernetics*, 103(4), 299-317.
- Dollé, L., Sheynikhovich, D., Girard, B., Ujfalussy, B., Chavarriaga, R., & Guillot, A. (2010b). Analyzing interactions between cue-guided and place-based navigation with a computational model of action selection: Influence of sensory cues and training. In *From Animals to Animals II* (pp. 335-346). Berlin-Heidelberg: Springer.
- Dollé L., Khamassi M., Girard B., Guillot A. & Chavarriaga R. (2008). *Analyzing interactions between navigation strategies using a computational model of action selection*. Berlin-Heidelberg:Springer, *Lecture Notes in Computer Science*, Spatial Cognition VI, 71-86.
- Duchon, A.P., Warren, W.H., Kaelbling, L.P. (1998). Ecological Robotics. *Adaptive Behavior*, 6(3/4), 473-507.
- Fikes R.E. & Nilsson N.J. (1971). STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. *Artificial Intelligence*, 2(3), 189-208.
- Firby, R.J. (1989). *Adaptive Execution in Complex Dynamic Worlds*. PhD Thesis, Yale University.

- Floreano, D., Ijspeert, A.J. & Schaal, S. (2014). Robotics and Neuroscience. *Current Biology*, 24(18), R910-R920.
- Foster, D.J., Morris, R.G.M., & Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus*, 10(1), 1-16.
- Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12), 493-501.
- Gat, E. (1991). Integrating reaction and planning in a heterogeneous asynchronous architecture for mobile robot navigation. *SIGART Bulletin*, 2(4), 70-74.
- Gibson, J.J. (1977). The theory of affordances. In R.E. Shaw & J. Bransford (eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies*, 144, 167-187.
- Girard B., Filliat D., Meyer J.A., Berthoz A. & Guillot A. (2005). Integration of navigation and action selection functionalities in a computational model of cortico-basal ganglia-thalamo-cortical loops. *Adaptive Behavior*, 13(2), 115-130.
- Glimcher, P.W. & Fehr, E. (eds.) (2013). *Neuroeconomics: Decision Making and the Brain*. New York: Academic Press.
- Goethe, J.W. von [1808]. *Faust. Eine Tragödie*. First published in 1808. Recent edition: Hamburger Lesehefte Verlag, Husum, 2010, ISBN 978-3-87291-028-8.
- Gray, J., Breazeal, C., Berlin, M., Brooks, A. & Lieberman, J. (2005). Action parsing and goal inference using self as simulator. In *IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005* (pp. 202-209). IEEE.
- Gray, J. & Breazeal, C. (2014). Manipulating mental states through physical action. *International Journal of Social Robotics*, 6(3), 315-327.
- Guazzelli, A., Bota, M., Corbacho, F. J., & Arbib, M. A. (1998). Affordances. Motivations, and the World Graph Theory. *Adaptive Behavior*, 6(3-4), 435-471.
- Hawkins, K.P., Bansal, S., Vo, N.N. & Bobick, A.F. (2014, May). Anticipating human actions for collaboration in the presence of task and sensor uncertainty. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 2215-2222). IEEE.
- Hasson, C. (2011). *Modélisation des mécanismes émotionnels pour un robot autonome : perspective développementale et sociale*. Thèse de doctorat. Université de Cergy Pontoise.
- Hertzberg, J., & Chatila, R. (2008). AI reasoning methods for robotics. In Siciliano, B. & Khatib, O. (eds.) *Handbook of Robotics* (pp. 207-223). Berlin: Springer.
- Hiatt, L. M., Harrison, A. M. & Trafton, J. G. (2011). Accommodating human variability in human-robot teams through theory of mind. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence* (Vol. 22, No. 3, p. 2066).
- Hiatt, L.M. & Trafton, J.G. (2010). A cognitive model of theory of mind. In *Proceedings of the 10th International Conference on Cognitive Modeling* (pp. 91-96).
- Ingrand, F., & Ghallab, M. (2015). Deliberation for autonomous robots: A survey. *Artificial Intelligence*, in press, corrected proof. Available online 28 November 2014.
- Jauffret, A., Cuperlier, N., Tarroux, P., & Gaussier, P. (2013). From self-assessment to frustration, a small step toward autonomy in robotic navigation. *Frontiers in neurobotics*, 7.
- Jeannerod, M. (1999). The 25th Bartlett Lecture. To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*, 52A, 1-29.

- Kanoun, O., Laumond, J. P., & Yoshida, E. (2011). Planning foot placements for a humanoid robot: A problem of inverse kinematics. *The International Journal of Robotics Research*, 30(4), 476-485.
- Keramati, M., Dezfouli, A., & Piray, P. (2011). Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS Computational Biology*, 7(5), e1002055.
- Khamassi, M., Enel, P., Dominey, P.F. & Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Progress in Brain Research*, 202, 441-464.
- Khamassi, M. & Humphries, M. D. (2012). Integrating cortico-limbic-basal ganglia architectures for learning model-based and model-free navigation strategies. *Frontiers in Behavioral Neuroscience*, 6:79.
- Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., & Guillot, A. (2005). Actor-Critic models of reinforcement learning in the basal ganglia: from natural to artificial rats. *Adaptive Behavior*, 13(2), 131-148.
- Khamassi, M., Lallée, S., Enel, P., Procyk, E. & Dominey, P.F. (2011). Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Frontiers in Neurobotics*, 5:1.
- Khamassi, M., Martinet, L. E., & Guillot, A. (2006). Combining self-organizing maps with mixtures of experts: application to an actor-critic model of reinforcement learning in the basal ganglia. In *From Animals to Animats 9* (pp. 394-405). Berlin Heidelberg: Springer.
- Khamassi, M., Wilson, C., Rothé, M., Quilodran, R., Dominey, P.F. & Procyk, E. (2011). Meta-learning, cognitive control, and physiological interactions between medial and lateral prefrontal cortex. In R.B. Mars, J. Sallet, M.F.S. Rushworth, & N. Yeung (eds.), *Neural Basis of Motivational and Cognitive Control* (pp. 351-370). Cambridge, Mass.: The MIT Press.
- Killcross, S. & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex*, 13(4), 400-408.
- Knoblich, G., Butterfill, S. & Sebanz, N. (2011). Psychological Research on Joint Action: Theory and Data. *Psychology of Learning and Motivation-Advances in Research and Theory*, 54, 59-101.
- Kober, J., & Peters, J. (2011). Learning elementary movements jointly with a higher level task. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on* (pp. 338-343). IEEE.
- Koechlin E., Ody C. & Kouneiher F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648), 1181-1115.
- Kortenkamp, D. & Simmons, R. (2008). Robotic systems architectures and programming. In Siciliano, B. & Khatib, O. (eds.) *Handbook of Robotics* (pp. 187-206). Berlin: Springer.
- Krichmar, J.L., & Edelman, G.M. (2002). Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cerebral Cortex*, 12(8), 818-830.
- Laumond, J.-P. (2012). *La robotique: une récidence d'Héphaïstos*. Paris: Éditions Fayard.
- Lehman, J.F., Laird, J. & Rosenbloom, P. (1996). A gentle introduction to soar: An architecture for human cognition. *Invitation to cognitive science*, 4, 212-249.
- Lemaignan, S., Ros, R., Sisbot, E.A., Alami, R. & Beetz, M. (2012). Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics*, 4(2), 181-199.
- Likhachev, M., Kaess, M. & Arkin, R. (2002). Learning behavioral parameterization using spatio-temporal case-based reasoning. In *Proceedings of IEEE International Conference on Robotics and Automation, ICRA '02*, (Vol. 2, pp. 1282-1289).

- Lungarella, M., Metta, G., Pfeifer, R. & Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, 15(4), 151-190.
- Mataric, M.J. & Michaud, F. (2008). Behavior-Based Systems. In Siciliano, B. & Khatib, O. (eds.) *Handbook of Robotics* (pp. 891-910). Berlin: Springer.
- McFarland, D. (1995). Opportunity versus Goals in Robots, Animals, and People. In Roitblat, H.L. & Meyer, J.-A. (eds.) *Comparative Approaches to Cognitive Science* (pp 415-434). Cambridge, Mass: The MIT Press.
- McGeer, T. (1990). Passive Dynamic Walking. *International Journal of Robotics Research*, 9(2), 62-82.
- Meyer, J.-A. (2015). *Dei Ex Machinis*, Les Éditions du Net.
- Meyer, J.A., & Guillot, A. (2008). Biologically inspired robots. In Springer Handbook of Robotics (pp. 1395-1422). Berlin-Heidelberg: Springer.
- Meyer, J.-A., Guillot, A., Girard, B., Khamassi, M., Pirim, P. & Berthoz, A. (2005). The Psikharpax project: towards building an artificial rat. *Robotics and Autonomous Systems*, 50(4), 211-223.
- Meyer, J.A. & Wilson, S.W. (1991). From Animals to Animats: Proc. 1st Intl. Conf. on Simulation of Adaptive Behavior. Cambridge, Mass.: The MIT Press.
- Miller, E.K. & Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Reviews in Neuroscience*, 24, 167-202.
- Milliez, G., Warnier, M., Clodic, A. & Alami, R. (2014). A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014 RO-MAN* (pp. 103-1109). IEEE.
- Minguez, J., Lamiroux, F. & Laumond, J.P. (2008). Motion planning and obstacle avoidance. In Siciliano, B. & Khatib, O. (eds.) *Handbook of Robotics* (pp. 827-852). Berlin: Springer.
- Morimoto, J., & Doya, K. (2001). Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *Robotics and Autonomous Systems*, 36(1), 37-51.
- Muscettola N., Nayak P.P., Pell B. & Williams B.C. (1998). Remote agent: To boldly go where no ai system has gone before. *Artificial Intelligence*, 103(1), 5-47.
- Mutlu, B., Terrell, A. & Huang, C.M. (2013). Coordination mechanisms in human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*.
- Nenas I., Wright A., Bajracharya M., Simmons R. & Estlin T. (2003). CLARAty and challenges of developing interoperable robotic software. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2003* (Vol. 3, pp. 2428-2435). IEEE.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, Mass.: Harvard University Press.
- Nikolaidis, S., & Shah, J. (2012). Human-robot teaming using shared mental models. In *ACM/IEEE HRI*.
- Nourbakhsh, I.R. & Fong, T. (2005). Human-robot teams on the moon: Challenges and plans. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, Workshop on Fielding Multi-Robot Systems.
- O'Keefe, J. & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford University Press.
- O'Regan, K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24, 939-973.
- Oudeyer, P.Y., Kaplan, F. & Hafner, V.V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2), 265-286.

- Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of learning and memory*, 65(1), 65-72.
- Pacherie E. (2012). Action. In K. Frankish & W. Ramsey (eds.), *The Cambridge Handbook of Cognitive Science* (pp. 92-111). Cambridge University Press.
- Peters, J. & Schaal, S. (2008). Reinforcement learning of motor skills with policy gradients. *Neural networks*, 21(4), 682-697.
- Peyrache A., Khamassi M., Benchenane K., Wiener S.I. & Battaglia F.P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, 12(7), 919-926.
- Pezzulo, G., Rigoli, F. & Chersi, F. (2013). The mixed instrumental controller: using value of information to combine habitual choice and mental simulation. *Frontiers in psychology*, 4.
- Pfeifer, R. & Pitti, A. (2012). *La révolution de l'intelligence du corps*. Paris: Manuella éditions.
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129-154.
- Redgrave, P., Prescott, T.J. & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4), 1009-1023.
- Renaudo, E., Girard, B., Chatila, R., & Khamassi, M. (2014). Design of a Control Architecture for Habit Learning in Robots. In *Biomimetic and Biohybrid Systems* (pp. 249-260). Berlin-Heidelberg: Springer.
- Renaudo, E., Girard, B., Chatila, R., & Khamassi, M. (2015a). Which criteria for autonomously shifting between goal-directed and habitual behaviors in robots? In *5th International Conference on Development and Learning and on Epigenetic Robotics (ICDL-EPIROB)* (Providence, RI).
- Renaudo, E., Girard, B., Chatila, R. & Khamassi, M. (2015b). Respective advantages and disadvantages of model-based and model-free reinforcement learning in a robotics neuro-inspired cognitive architecture. In *Proceedings of Biologically Inspired Cognitive Architectures BICA 2015*, Lyon, France. Procedia Computer Science (Elsevier).
- Renaudo, E., Devin, S., Girard, B., Chatila, R., Alami, R., Khamassi, M. & Clodic, A. (2015c). Learning to interact with humans using goal-directed and habitual behaviors. *Workshop on Learning for Human-Robot Collaboration at RO-MAN 2015 Conference*. Kobe, Japan.
- Resulaj, A., Kiani, R., Wolpert, D.M. & Shadlen, M.N. (2009). Changes of mind in decision-making. *Nature*, 461, 263-266.
- Richardson, M.J., Marsh, K.L., Isenhower, R.W., Goodman, J.R.L. & Schmidt, R.C. (2007). Rocking together: Dynamics of unintentional and intentional interpersonal coordination. *Human Movement Science*, 26, 867-891.
- Rizzolatti, G. & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11, 264-274.
- Roitblat, H.L. & Meyer, J.-A. (eds.) (1995). *Comparative Approaches to Cognitive Science* (Introduction, pp. 1-9). Cambridge, Mass.: The MIT Press.
- Rolf, M., Steil, J.J. & Gienger, M. (2011). Online goal babbling for rapid bootstrapping of inverse models in high dimensions. In *Proceedings of the IEEE International Conference on Development and Learning (ICDL)* (Vol. 2, pp. 1-8). IEEE.
- Rondi-Reig, L., Petit, G. H., Tobin, C., Tonegawa, S., Mariani, J., & Berthoz, A. (2006). Impaired sequential egocentric and allocentric memories in forebrain-specific-NMDA receptor knock-out mice during a new task dissociating strategies of navigation. *The Journal of Neuroscience*, 26(15), 4071-4081.
- Rosenbloom P., Laird J. & Newell A. (Eds.) (1993). *The Soar Papers: Research on Integrated Intelligence*. Cambridge, Mass.: The MIT Press.

- Santucci, V.G., Baldassarre, G. & Mirolli, M. (2016). GRAIL: a Goal-discovering Robotic Architecture for Intrinsically-motivated Learning. *IEEE Transactions on Autonomous Mental Development*. In press
- Saridis, G.N. (1995). Architectures for intelligent controls. *Intelligent control systems: Theory and applications*. Piscataway, NJ.: IEEE Press,
- Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2), 173-187.
- Shadmehr, R. & Wise, S.P. (2005). *Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. Cambridge, Mass: The MIT Press.
- Shors, T.J. (2004). Memory traces of trace memories: neurogenesis, synaptogenesis and awareness. *Trends Neuroscience*, 27(5), 250-256.
- Siciliano, B. & Khatib, O. (eds.) (2008). *Springer Handbook of Robotics*. Berlin-Heidelberg: Springer.
- Sigaud, O., & Peters, J. (eds.) (2010). *From Motor Learning to Interaction Learning in Robots*. Berlin-Heidelberg: Springer.
- Sisbot, E.A., Ros, R. & Alami, R. (2011). Situation assessment for human-robot interactive object manipulation. In *RO-MAN, 2011 IEEE* (pp. 15-20). IEEE.
- Smart, W.D., & Kaelbling, L.P. (2002). Effective reinforcement learning for mobile robots. In *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on* (Vol. 4, pp. 3404-3410). IEEE.
- Smith, E.E. & Kosslyn, S.M. (2007). *Cognitive Psychology: Mind and Brain (1st ed.)*. Upper Saddle River, NJ.: Pearson/Prentice Hall.
- Staudte, M. & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, 120(2), 268-291.
- Stulp, F., & Sigaud, O. (2013). Robot skill learning: From reinforcement learning to evolution strategies. *Paladyn, Journal of Behavioral Robotics*, 4(1), 49-61.
- Tambe M. (1997). Towards flexible teamwork. *Journal of Artificial Intelligence Research*, 7, 83-124.
- Toates, F. (1995). Animal Motivation and Cognition. In Roitblat, H.L. & Meyer, J.-A. (eds.) *Comparative Approaches to Cognitive Science* (pp. 435-464). Cambridge, Mass.: The MIT Press.
- Tomasello, M. & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121-125.
- Tuomela, R. (2013). *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press.
- Ullman, M.T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- van Schie, H.T., Mars, R.B., Coles, M.G. & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, 7(5), 549-554.
- van Ulzen, N.R., Lamoth, C.J., Daffertshofer, A., Semin, G.R. & Beek, P.J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience letters*, 432(2), 88-93.
- Varela, F.J., Thompson, E., & Rosch, E. (1993). *The Embodied Mind*. Cambridge, Mass.: The MIT Press.
- Vernon, D. (2014). *Artificial Cognitive Systems - A Primer*. Cambridge, Mass.: The MIT Press.
- Vesper, C. (2014). How to support action prediction: Evidence from human coordination tasks. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 2014 RO-MAN* (pp. 655-659). IEEE.
- Viejo, G., Khamassi, M., Brovelli, A., & Girard, B. (2015). Modeling choice and reaction time during arbitrary visuomotor learning through the coordination of adaptive working memory and reinforcement learning. *Frontiers in Behavioral Neuroscience*, 9.

- Volpe, R., Nesnas, I.A.D., Estlin, T., Mutz, D., Petras, R. & Das, H. (2001). The clarity architecture for robotic autonomy. In *IEEE 2001 Aerospace Conference*.
- Warnier, J.-P. (2010). Physiologie de l'action et culture matérielle. *Intellectica*, 53-54, 181-194.
- Weng, J. (2004). Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, 1(02), 199-236.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460.
- Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.
- Yin, H.H. & Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6), 464-476.