Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

NEUROSCIENZE COGNITIVE

Ciclo XXIV

Settore Concorsuale di afferenza: AREA 11
Settore scientifico-disciplinare di afferenza: M-PSI/02

SOCIAL LEARNING AND ACTION UNDERSTANDING IN HUMAN OBSERVERS: CONTRIBUTIONS OF SENSORI-MOTOR CONSTRAINTS AND PRIOR INFORMATION

Presentata da: PIERRE OLIVIER JACQUET

Coordinatore Dottorato	**Relatore**
ELISABETTA LADAVAS	**ANNA MARIA BORGHI**

Esame finale anno 2012

Acknowledgments

First of all, I am deeply grateful to Professor Anna Maria Borghi, who gave me the very first chance to become a researcher. I warmly thank her for all what she taught to me, and, most of all, for her honesty, openness, kindness, and her benevolent patience.

I owe Doctor Alessandro Farnè, Doctor Karen Reilly, and Doctor Alice Roy, a huge debt of gratitude as they welcomed me in their laboratory in Lyon, and simply because they are extraordinary people. They teached me so much, and believed in me and in my work much more than I do myself.

I would like to address a special thanks to Doctor Karen Reilly – my big sister – for having supported me, and for having endured my moods, my strain and stress, since the very beginning. I thank her for having helped me for hours by phone, by emails, and even by telepathy – who knows. But most of all, I thank her for having introduced me to Neutral Milk Hotel.

I thank Professor Alessia Tessari who has believed in me and who motivated me to get a job of Research Assistant for the European project ROSSI (Emergence of communication in RObots through Sensorimotor and Social Interaction), at the Bologna University, where everything began.

I also thank the team of the Center for the Study and Research in Cognitive Neuroscience of the Bologna University, and especially Doctors Alessio Avenanti, Andrea Serino, Professors Giuseppe di Pellegrino and Elisabetta Ladavas, who gave me the opportunity to do this PhD program.

I also would like to thank all members of the European project ROSSI, all members of the EMCo Lab of the Bologna University, and all members of the ImpAct Team at the INSERM unit in Lyon, with a special thought to Claudia Scorolli who describes the Simon Effect better than anybody else, and to Stéphane Jacobs who shares my passion for the one-handed backhand.

I deeply thank Valerian 'Slash' Chambon for, well, everything, and much more...I thank Alberto Acerbi for the Fernet-Branca, 'il Banana' and 'la Piscina', and Stéphanie Bellocchi for having endured the Fernet-Branca, 'il Banana' and 'la Piscina'. My thoughts especially go to Jérôme Munuera with whom I should be training my footwork in order to represent an unlikely flag at the London 2012 Olympic Games, instead of writing these lines. I also thank Matthieu and Julie, my Alain, Karine and my nephew Lilou, Perrililia, Lolo 'the Freedom Man' and Karine, Joce & Pat, Thomas 'the Juggler' Castelain, Nabou, JVDW, Richard Brautigan, Doon, Hélène, Nance, Zazou, Pedro, Jacques Lambert, Stoïc/Monique, Kap Bambino, Bénoche, the Baumard's 'Garde la Morale' Family, Julie Bonneric, Guy Perrillat, Richard Dawkins, Bibi, 'Pistol' Pete Sampras, my italian teacher Franco, Teen Days Running, Hugo, Kara Thrace, Gaius Baltar, Claudia G., Neutral Milk Hotel, Flavia, Thônes Represent, Caro Maire, Dan Dennett, Laura, Domenech, Luke, Elvis, Mathzou, Le Piou, Roger Federer, Alain & Françoise Combette, Nicolas Franck, Rob Crow, Tonton Tourny, Kim Sterelny, The Gerelli's Family and Friends, and many others that I probably forget. All these people inspire me so much.

I also thank ALL my family, but especially: my Mum for having brought me into the World – and many many more –, my Grand-Parents André and Francine for almost the same reasons; my Dad and Monique for never asking me what the hell I was doing with my life all these years; my sister Delphine and her family, my sister Bea for all the things she did for me since a very long time; my cousin/brother Nano and her girlfriend Carolus; and a deep, and eternal thought to my cousin/brother Stéphane and his girlfriend Laurence.

Finally, I thank Aline, the one with whom I share my life, the one with whom I laugh, the one with whom I can cry, the one with whom everything is possible, the one that makes me better, the one who makes me crazy, the one with whom I am having a baby, the one I am in love with.

General abstract

The general aim of this thesis was to investigate the respective contribution of prior information and sensorimotor constraints to action understanding and prediction, and to estimate their consequences on the evolution of faithful social learning. Even though a huge amount of literature has been dedicated to the study of action understanding and its role in social learning, these issues are still largely debated. Here, I critically describe two main perspectives. The first perspective (i.e., the 'Theory-theory' framework) interprets faithful social learning as an outcome of a fine-grained representation of others' actions and intentions that requires sophisticated – and uniquely human – socio-cognitive skills. In contrast, the second perspective (i.e., the 'simulation theory' framework) highlights the role of simpler decision heuristics, the recruitment of which is determined by individual and ecological constraints. The aim of the present thesis is to provide evidence that these two theoretical contributions are not mutually exclusive.

The general introduction is organised around five sections in which I address the following points: i) the relationship between social learning and theories of action understanding; ii) the role of prior information in solving the inverse problem addressed by action understanding; iii) the role of rationality in inference mechanisms involved in action understanding; iv) the role of object affordances; v) the hybrid model of action understanding.

The experimental contribution of the thesis is divided into four main studies:

- The first study aimed to investigate the role of the inferior frontal cortex (IFC), the anterior intraparietal area (AIP) and the primary somatosensory cortex (S1) in the recognition of other people's actions, using a transcranial magnetic stimulation adaptation paradigm (TMSA). In a first adaptation phase, participants were repeatedly exposed to goal-directed actions performed by a filmed demonstrator. In a second recognition phase, they were presented pictures of actions and were asked to match either the type grip or the type of arm action currently presented with those used in the preceding adaptation phase. Single-pulse TMS was applied over the three regions of interest during the presentation of each picture. Overall, results showed that stimulating the IFC and S1 improved the recognition of the adapted arm action, possibly through of 'simulation' process of sensorimotor and somatic properties of the observed actions.
- The second work aimed at studying whether, and how, prior information acquired from the probabilistic sampling of past events and prior information derived from an estimation of sensorimotor/biomechanical constraints of observed goal-directed actions interact during the prediction of other people's intentions. Participants performed an action prediction task in which they were required to infer, under various conditions of visual uncertainty, the intentions of a demonstrator performing tool-use behaviors. Both the probability of observing the demonstrator achieving a particular tool function and the biomechanical optimality of the observed movement were varied. Results showed that biomechanical priors modulate the extent to which participants' predictions are influenced by probabilistically-induced expectations.
- In line with this finding, the third study aimed to investigate whether, and how, the interactions between probabilistic and biomechanical priors modulate motor system activity. This modulation was measured by means of single-pulse TMS applied over the primary motor cortex (M1) during action prediction under visual uncertainty, before and after probabilistic exposure to biomechanically optimal and suboptimal actions. Results revealed a suppression of the corticospinal excitability, whose magnitude was found to depend on the type of actions (biomechanically optimal or suboptimal) that were probabilistically biased.
- The fourth study tested the extent to which behavioral and ecological constraints – such as the richness of individuals' biomechanical repertoire or the type of search space characterizing a problem – influence at a population level the emergence of faithful social learning strategies (e.g., emulation or imitation), and whether the emergence of such strategies may explain the stability of observed behavioral traditions. These relationships where explored through an evolutionary individual-based model. Results showed that the evolution of faithful social learning only occurs if the behavioral repertoire of a population is large enough, and if the search space does not allow trial-and-error learning.

Overall, the collected data contribute to our understanding of action prediction by humans, by elucidating how higher-order and lower-order prior expectations interact during action prediction. In addition, these studies offer a new point of view on the neural underpinnings of action representation and action prediction. Finally, this work provides promising perspectives for a better understanding of human social learning, with possible extensions to animal models.

Table of contents

GENERAL INTRODUCTION

The present thesis focuses on the cognitive and cerebral mechanisms underlying how human observers perceive, predict and learn from their conspecific's behaviors. It also aims to extend the result of such investigations to a broader level of description, such as the emergence of cultural traditions in a population of individuals.

Understanding human culture requires the investigation of at least three main aspects: the population-level patterning of traditions; the intermediate level of social learning mechanisms; and the individual-level behavioural and cognitive characteristics (Whiten, 2011). The present thesis contributions to each of these aspects, with particular attention devoted to the small-scale, individual level. Throughout these pages the reader will thus encounter theoretical and experimental work that, put together, aims to draw links between the individual-level and the larger levels of description mentioned above. The main claim of the present manuscript stems from the well accepted assumption that perceiving and understanding the behaviors of their conspecifics through the situations in which they take place is the precondition for social learning and the emergence of culture (Sperber, 1996; Sperber and Hirschfeld, 2004).

Traditionally, it has been claimed that the emergence of cultural traditions depends on the use of faithful social learning (i.e., the transmission of information from an individual to another through behavioral means, in particular, through learning and teaching), that faithful social learning is an outcome of a fine-grained representation of others' actions and intentions, and that this representation requires sophisticated – and uniquely human – socio-cognitive skills. In the present thesis, in contrast, I suggest a less exclusive and more integrative position. This position proposes that a fine-grained representation of other people's actions and intentions requires sophisticated socio-cognitive skills as well as simpler heuristics and decision-making rules, whose recruitment is determined by individual and ecological constraints. These constraints may have a significant impact on the evolution of social-learning and, thus, on the emergence of cultural traditions.

The present work is divided into three main parts:

a) A general introduction that aims to identify the sophisticated socio-cognitive skills and simple heuristics that preside over human action understanding, and that underlines their cerebral correlates.

b) A general method section in which the experimental techniques used in the presented studies are described.

c) An experimental part that aims:
   i. to study the cerebral bases of action perception
   ii. to investigate whether, and how, sophisticated skills and simple heuristics interact during action prediction
   iii. to investigate how this interaction modulates brain activity
   iv. to investigate whether, and how, individual and ecological constraints can impact on the emergence of ‘cultural’ traditions among a population.

d) A general discussion where experimental data are discussed in the larger context of human social learning.

A. From social learning to action understanding

Behavioral – or cultural – traditions (i.e., behaviors acquired through some form of social learning, and which are relatively stable in groups) have been found in humans and in several non-human species (Laland and Galef, 2009). The existence of these traditions has usually been considered as a strong marker of faithful social transmission mechanisms that guarantee both the successful diffusion and the stability of the behaviors (Huffman, 1996; Horner et al., 2006; Marino et al., 2007). Social learning – that is, the 'inheritance of acquired behaviors' – has been a longstanding interest of biologists and psychologists alike. Indeed, the idea that animals acquire components of their behavioral repertoire by copying their conspecifics is far from being new, as witnessed by seminal observations made by Aristotle, for whom human and non-human social animals have, in varying degrees, natural predispositions to interindividual communication and social learning (1986). The social learning of new skills, ranging from the acquisition of complex tool use by humans to the acquisition of vocalization by birds, is now considered as a well-established and noncontroversial aspect of the adaptive behavior of vertebrates (Seed and Byrne, 2010). However, if a great deal is known about the adaptive functions of social learning, such as its impact on the social and cultural transmission of behaviors, little is known about the cognitive mechanisms that make them possible.

Many forms of social learning mechanisms operate in human behaviors. These forms are hierarchically organized according to their efficiency for transmitting the behavior to a third party with fidelity (Rendell et al., 2011). A distinction is usually made between:

- Stimulus enhancement: a demonstrator exposes an observer to a single stimulus, which leads to a change in the probability that the observer will respond to stimuli of that type.
- Local enhancement: a demonstrator attracts an observer to a specific location, which can lead to the observer learning about objects at that location.
- Observational conditioning: the behaviour of the demonstrator exposes an observer to a relationship between stimuli, enabling the observer to form an association between them.
- Response facilitation: a demonstrator performing an action increases the probability that an observer will do the same. This can result in the observer learning about the context in which to perform the action and the consequences of doing so.
- Social facilitation: social facilitation occurs when the mere presence of a demonstrator affects the observer's behavior, which can influence the observer's learning.
- Contextual imitation: observing a demonstrator performing an action in a specific context directly improves learning about how to perform this action in the same context.
- Production imitation: observing a demonstrator performing a novel action, or novel action sequence, increases the likelihood that an observer performs that action or sequence.
- Emulation: observation of a demonstrator interacting with objects in its environment increases the likelihood that an observer performs any actions that bring about a similar effect on those objects.

Local enhancement, observational conditioning, response facilitation and social facilitation are forms of incidental, low-fidelity social learning and are most commonly observed in non-human animals. They involve the detection of contextual information, of low-level social signals (e.g., the mere presence of a peer) or basic reinforcement signals. These forms of social-learning are not directly ‘social’. Instead, they result from the fact that individuals live in a social environment. In these cases, the learned information is mediated through a social channel (Sterelny, 2009) – i.e., through the observation of others’ behavior – but the cognitive processes that encode this information are just the same as those that encode information received from other channels. They are not adaptively specialized in any way for the processing of socially-channeled information.

In contrast, emulation, contextual imitation, and production imitation are forms of motivated, faithful (high-fidelity) social learning that have been rarely reported in non-human animals, except in primates (Voelkl and Huber, 2007) and birds (Akins and Zentall, 1998; Dorrance and Zentall, 2001). Among primates, humans are well known to be precocious and efficient imitators and emulators (Tomasello, 1999). The relative rarity of emulation and imitation in the wild suggests that these sophisticated forms of social learning are the hallmark of a cognitive specialization for processing social signals (Tomasello et al., 2005). These faithful forms of social-learning are thus patently ‘social’ because they require the observer to be intrinsically motivated in directing her/his attention towards the behavior of the demonstrator. This is particularly salient with humans who are equipped with a set of psychological dispositions biasing the individual to preferentially orient to the social world (Chevallier et al., 2012; Csibra and Gergely, 2011).

Evolving culture through high-fidelity social transmission mechanisms does not merely rely on social motivation. For example, imitation refers to the ability of overtly copying an observed action, while emulation refers to the ability to copy the outcome of an observed action through the observer’s own means. Thus, copying with fidelity a new behavior from mere observation has been tightly related to the observer’s ability to build a fine-grained representation of the observed motor sequence, and/or a representation of its underlying cause (the goal or the intention of the observed agent). In sum, it requires a deep understanding of other people’s behaviors and the mental states that cause these behaviors (Csibra and Gergely, 2007).

1. Interpreting other people’s behaviors: from theory-theory to motor simulation

Perceiving and understanding our conspecific’s behaviors is one of the pre-requisites for the emergence of human culture (Tomasello et al., 2005). The typically human disposition to understand and predict others’ behaviors render an account for a more fundamental aptitude to represent, or ’read’, other people’s mental states. Understanding others’ behaviors would thus depend on this robust and early ability to represent and attribute mental states (Baron-Cohen et al., 1985; Leslie, 1987; Frith and Frith, 2003). Yet, the exact nature of the elementary mechanisms this ‘mind-reading’ ability relies on is still a matter of debate.

1.1. The ‘theory-theory’

According to the advocates of the ‘Theory-theory’ framework, understanding behaviors of our peers would require a preliminary, implicit and tacit ‘naive’ theory of psychological functioning, involving axioms and elementary inferential mechanisms (Gopnik, 1993; Gopnik and Meltzoff, 1994). These axioms and inferential mechanisms would be, at least partly (Carruthers, 1996), based on past experience of the observer, and on laws extracted from the regularity of past events (Leslie, 1987; Gopnik, 1993). Just like other naive theories, such as naive physics, the naive theory of psychological functioning enables us to predict and explain another person’s behavior (Saxe, 2005). Although we daily use this theory, we are not actually aware of its underlying laws.

The ‘Theory-theory’ approach posits that the emergence of our capacity to explain others’ behaviors in terms of psychological states – such as ‘beliefs’, ‘desires’ or ‘intentions’ – intimately depends on interactions experienced by the child with her/his relatives. These interactions progressively lead the child to form hypotheses about hidden variables (e.g., the beliefs and desires of her/his relatives) that may explain the behavioral regularities she/he observes.According to Gopnik and Meltzoff (1994), the child would resemble a ‘little scientist’ testing successive hypotheses about what may have caused the observed behaviors (‘he believes that’ or ‘he wants that’). Testing these hypotheses would lead the child to progressively elaborate a ‘naive’ theory of the psychological functioning of her/his conspecifics. At a very early stage, the child would possess a primitive concept of ‘belief’ and ‘desire’ (the child’s beliefs are always ‘true’ and other people always express the same desires as her/his own) that would be refined with experience and development. Thus, the ability to attribute intentions or beliefs to other people would be related to the elementary ability to infer mental states causing the observed behaviors and the situation, or context, in which such behaviors take place. Interestingly, these inferential rules would be crucial to understand others as well as to build knowledge about oneself. Thus, self-knowledge itself would not be immediate, but inferential, i.e., mediated by automatic inferences achieved in the presence of biological agents. Nonetheless, a difference of degree would exist between knowing oneself and knowing others. Indeed, information we have about ourselves would be more abundant, and of a better quality, as we naturally are in a better position to collect it. However, this difference of degree would not be a difference of nature. In one case (knowing oneself) as in the other (knowing others), we would use the same implicit theory of mind functioning (Gopnik and Meltzoff (1994).

1.2. Pitfalls of the Theory-theory: the theory of simulation

A strong argument against advocates of the "theory-theory" is that the formation of a folk psychology seems to be a relatively late developmental achievement: the ability to attribute 'false beliefs', for example, only emerges around 4 years of age (Wimer and Perner, 1983). Recent data collected in the field of neuroscience – and more particularly in the field of motor cognition – provides some convincing counterarguments to the 'Theory-theory' approach (see for a review, Gallese and Goldmann, 1998). According to these studies, the perception of others' mental states would not require the acquisition of a 'model' of the other's mind. Instead, humans already possess such a model, namely, their own mind. Thus, understanding other people's behavior would primarily depend on the observer's ability to 'simulate' the ‘point of view’ of the observed agent (Goldmann, 1995; Gordon, 1996; Gallese and Goldmann, 1998).

The acquisition by humans of mental concepts, such as understanding and predicting intentional behaviors, would depend on this ability to internally replicate (i.e., ‘simulate’) other people’s behaviors. Here, the simulation process would be based on a representation of sensori-motor constraints that are shared between the observer and the demonstrator. As the observer may form motor representations that are similar to those of the observed agent, she would also be able to (automatically) access the intentions underlying, and causing, such representations (Wolpert et al., 2003; Blakemore & Decety, 2001; Metzinger & Gallese, 2003).

1.2.1. A brief historical account of simulationist theories

Simulationist theory is divided into two main approaches: the introspectionist approach and the anti-introspectionist approach. The introspectionnist approach has been mainly popularized by Goldman (1995), according to whom simulating an observed behavior requires prior knowledge about what ‘believing’ or ‘knowing’ means. Such knowledge is acquired through introspection and is the necessary precondition for a simulation process to be effective. In this introspectionist perspective, the simulation process is ‘attributor-dependent’, meaning that the observer is the reference for the simulation. We often put ourselves into the situation of people we observe, and approximate what they might think, but the simulation in itself does not elicit in the simulator the very same states as those experienced by the simulated target.

In contrast, the anti-introspectionist approach developed by Gordon (1996) posits that the simulation process is ‘attributor-neutral’, meaning that simulation is a transformation rather than a transfer. Simulation is the transformation of one’s own situation into the other’s situation. Here, the introspective dimension is excluded from the simulation. It consists of simulating a behavior without using any concepts of folk psychology such as ‘belief’ or ‘desire’. Rather, it consists of experiencing the values or sensations associated with the simulated behavior. For example, a child would not simulate her/his relatives as if they had specific psychological properties (e.g., belief, desire or intention). Instead, when a child faces a particular situation involving the behavior of a relative, s/he reconstructs the value of the observed behavior by simulating it, and concludes that the situation has certain properties: motivational, emotional, etc. In this case, the target of the simulation is more the context in which an action takes place than the psychological dispositions of the agent. The context immediately elicits relational properties about the interest or the danger of a behavior achieved in such a context.

1.2.2. The theory of motor simulation

The existence of this interpretative system based on shared representations has been supported by the discovery, in monkeys and more recently in birds, of a new type of visuomotor neuron – i.e., the so-called ‘mirror neurons’. These bimodal neurons were originally found in the ventral premotor cortex of macaques and have been shown to discharge both when the animal performs an action and sees the same action performed by a third party (di Pellegrino et al., 1992; Rizzolatti et al., 1996; Gallese et al., 1996; Rizzolatti et al., 2000). Mirror neuron activity has been mainly recorded during the observation of specific goal-related motor acts (such as reaching and grasping). The firing of these ‘mirror’ neurons would *simulate,* in the observer’s brain, the observed movements of the demonstrator: thus, in the absence of overt motor execution, a part of the observer’s motor system is active as if they were actually executing this action. That is, the brain transforms the visual information conveyed by the observed action’s kinematics into equivalent (non-executed) motor commands. Rizzolatti and Arbib (1998) described the functional role of mirror neurons as follows:

> ‘The response properties of mirror neurons to visual stimuli can be summarized as follows: mirror neurons do not discharge in response to object presentation; in order to be triggered they require a specific observed action. The majority of them respond selectively when the monkey observes one type of action (such as grasping). Some are highly specific, coding not only the action aim, but also how that action is executed. They fire, for example, during observation of grasping movements, but only when the object is grasped with the index finger and the thumb.’ (p. 188)

The automatic matching of the demonstrator’s movements with motor representations stored in the observer’s behavioral repertoire would enable the formation of shared motor representations (Jeannerod, 2001). In line with this claim, it has been suggested that mirror neurons are involved in various cognitive domains, ranging from imitation to language acquisition (Rizzolatti and Arbib, 1998). Furthermore, Gallese & Goldman (1998) suggest that one function of mirror neurons is to ‘enable an organism to detect certain mental states of observed conspecifics. This function may be part of, or a precursor to a more general mind-reading ability’ (p. 493). Thus, the activation of these shared motor representations would allow the observer to access the goals and intentions of the *simulated* demonstrator and, by extension, to draw predictions regarding her/his future behaviors.

The presence in the monkey brain of visuomotor neurons endowed with ‘mirror’ properties is well documented and generally agreed upon. However, their presence in the human brain is much more controversial (Hickok, 2009), as the correlational techniques used in human neuroscience provide no direct evidence for their existence (but see Mukamel et al., 2010). Several studies using functional neuroimaging have, however, suggested the existence of a cortical network (see figure 1) that is common to the observation, the imagination, and the execution of a specific action (Iacoboni et al., 1999; Rizzolatti et al., 2001 ; Grèzes et al., 2003). This network, called the mirror system or the Action Observation Network (AON), is distributed along an antero-posterior axis that classicaly involves several cerebral structures such as:

- the superior temporal sulcus (STS) that, in its posterior part, has multimodal integration properties (Barnes and Pandya, 1992), is recruited during the observation of movements of biological agents (Puce and Perrett, 2003). However, an increasing number of studies suggests that the STS also responds to the relationship between an observed movement and the structure of the environment, and code for the observation of goal-directed movements (German et al., 2004; Saxe et al., 2004)
- the inferior parietal lobule (IPL) and, more specifically, the anterior intraparietal sulcus (AIPs) that is involved in the visual guidance of grasping movements (Faillenot et al., 1997) as well as in the decoding of observed movements (Bonda et al., 1996). It

has also been shown that AIPs was sensitive to the goals of actions (Hamiltion and Grafton, 2006).

- the ventral part of the premotor cortex (PMv) and, more generally, the inferior frontal cortex (IFC), known to be involved in the execution as well as the observation of intentional, goal-directed actions (Grèzes and Decéty, 2001; Iacoboni et al., 1999; Iacoboni et al., 2005; Buccino et al., 2001)

Of note is that the AON may involve other sensory regions that are not classicaly reported as playing a role in action perception and understanding. In particular, the somatosensory cortices may be recruited during the observation of actions and may play a role in the estimation of their sensory consequences (Keysers and Gazzola, 2007, 2010; Kilner, 2011).

According to the most popular assumption, known as *the direct-matching hypothesis*, the AON underlies the processes of imitation and intention attribution. Indeed, the direct-matching hypothesis presupposes that ‘an action is understood when its observation causes the motor system to resonate’ (Rizzolatti et al., 2001). This motor *resonance* would allow an observer to figure out the outcome of the perceived action, for the observer ‘knows its outcome when he does it’ (Gallese et al., 2004). Thus, mirroring an observed action would provide a simulation device for goal understanding by replicating the observed action in the observer’s own motor repertoire. Accordingly, the represented action should be ‘executable’, that is, consistent with the constraints of the observer’s motor system.

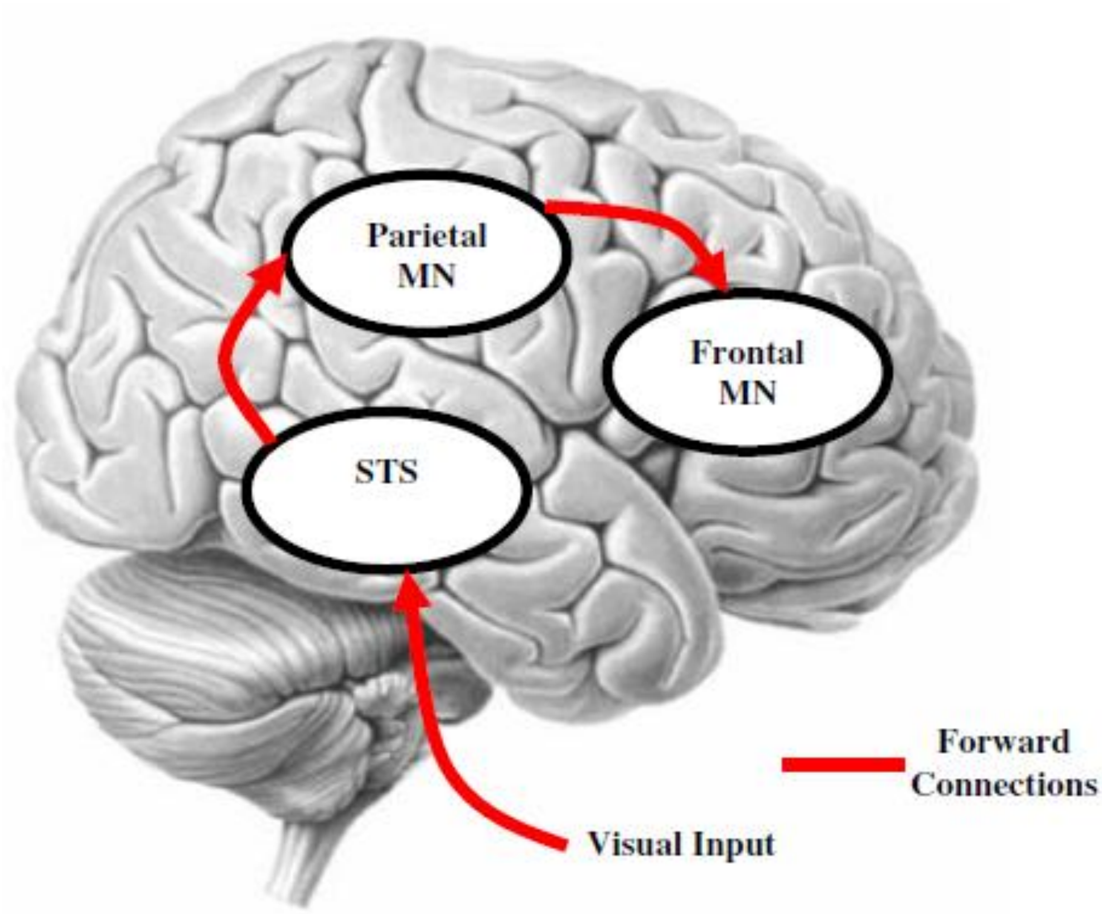


Figure 1. Representation of the human 'mirror system', also called Action Observation Network (taken from Kilner et al., 2008). The observation of an action drives the firing of neurons in the superior temporal sulcus (STS), which drives activity in the inferior parietal area, which in turn drives activity in inferior frontal gyrus.

However, a specific movement – let say, a grasping movement – may have been executed for several different reasons. And the decoding of the kinematics only would not be informative enough to allow the observer to unambiguously select the intention that has caused the observed action. Recent studies have demonstrated that the predictive function of mirror neurons was varying according to the context in which the observed action scene is embedded (Iacoboni et al., 2005; Fogassi et al., 2005; Chersi et al., 2011). More specifically, the selection of an appropriate intention would be done through a process of action reconstruction. Fogassi and co-workers (2005) proposed that the observation of a motor act that is at the origins of a given action would trigger within the observer's mirror system the subsequent motor act that is likely to be associated with the former, and so on until the action chain is achieved. Accorrding to these authors, such reconstruction process of action chains would occur within the parietal lobe in which motor acts would not be represented independently of the global aim of the action, but rather, would form prewired intentional chains in which each motor act is facilitated by the previously executed one. Consequently, the activation of a specific action chain within the mirror system would allow the observer to form a representation of the intention that, most likely, the agent is about to achieve. Interestingly, Iacoboni and colleagues (2005) suggested that the activation of an action chain was specified by contextual information processed in cerebral regions that are reciprocally connected to the mirror system. This contextual information can be the situational constraints in which the action takes place (e.g., if the scene provides additional non-motor cues about which intention the observed agent is more likely to achieve), or the affordances of objects (e.g., their size, shape, density or texture) that are the target of an action.

In summary, these mechanisms of shared activations of motor representations provide convincing arguments in favor of the simulationist theory. Our natural capacities to simulate – that are underpinned by the existence of 'mirror systems' of shared representations – would provide access to the mental states of others – such as the motor intentions that guide their actions – without the need for any naive theory about human psychology or about mind functioning in general (Blakemore and Decéty, 2001 ; Gallese, 2003 ; Keysers and Gazzola, 2007). Moreover, the theory of motor simulation resolves a developmental problem: indeed, how does theory-theory explain the early capacity of human infants to interpret other people's goal-directed behaviors, as young infants do not yet possess psychological concepts such as 'belief', 'desire', or 'intention'? Accordingly, motor simulation, based on a hard-wired property of the motor system, itself, is a good candidate to explain the late elaboration of more complex interpretative strategies (such as an inferential theory of action understanding).

However, the exact functional role of mirror neurons in action understanding and action prediction remains debated (Jacob and Jeannerod, 2005; Saxe, 2005; Hickok, 2009; Mukamel et al., 2010; Rizzolati and Craighero, 2004; Rizzolatti et al., 2001). It has been suggested that their contribution to the domain of action understanding depends on the level of complexity of the observed behavior. Indeed, an observed action can be 'understood' at multiple levels, depending on its temporal and structural complexity.

2. Hierarchical representation of actions in the brain

Recent advances in human neuroscience suggest that complex behaviors are hierarchically organized around distal outcomes (Jeannerod, 2004; Hamilton and Grafton, 2007). Let's take the example of an observer watching another agent grasping and lifting a glass full of red wine. The observer will roughly decompose this action into a hierarchically organized chain of several motor sub-steps, each ultimately achieving different sub-goals. At the top of the hierarchy is the agent's higher-level goal, which is to lift the glass once it has been grasped. The overarching goal requires the achievement of a sequence of distinct motor steps. One of these sub-steps may be to grasp the glass with the whole hand, which can be considered as a sub-goal to reach the higher-level goal. But in order to achieve this sub-goal, the agent has to perform a series of motor acts that, in turn, can be decomposed into finer motor patterns (e.g., activating hand muscles during the opening phase of the grasping movement). Of note is that the higher-level goal can be achieved by the observed agent using a number of alternative sequences of motor acts. For example, instead of grasping the glass with the whole hand, the agent may also try to grasp it with two-fingers, or even with the mouth. Although the aim of these alternatives is same higher-level goal as the one first described, they recruit a different set of motor sub-steps. In the same way, an identical chain of motor sub-steps (e.g., reaching, grasping and lifting a glass full of red wine) can lead to different final intentions (e.g., to toast somebody versus to check the colour of the wine).

Actions can thus be roughly organized according to four levels (Kilner, 2011):

i) the kinematic level: the trajectory and velocity of the action, including both the reach and grasp phase of a goal-directed action

ii) the motor level: the pattern of muscle activity required to produce the kinematics

iii) the goal level: the immediate purpose of the action, such as to grasp an object

iv) the intention level: the overarching cause – or reason – of the action

These four levels are not independent, but hierarchically organized: the kinematic level is dependent on the motor level; the motor level is dependent on the goal level, and the goal level is in turn dependent on the intention level. This hierarchy also defines a continuum of abstraction, with the intention level defining the most abstract features of action representation. For example, a specific intention (e.g., pouring liquid into a glass) can be achieved through many different motor acts: as such, there is no univocal (i.e., one-to-one) mapping between the content of the intention and the set of motor alternatives that can potentially achieve it.

These different levels of representation of goal-directed actions have been shown to be distributed within a large cerebral network:

1. the posterior parietal cortex (PPC) is involved in the processing of the lower level, concrete features of the action (i.e., the kinematic level) and the intentional planning of movements (Andersen and Buneo, 2002).
2. the anterior intra-parietal sulcus (AIPs) is known to be involved in the processing of goal-directed prehension movements, and is assumed to code perceived interactions between objects and goals (Faillenot et al., 1997; Hamilton and Grafton, 2006)
3. the inferior parietal lobule (IPL) and
4. the inferior frontal gyrus (IFG) would both be involved in the processing of the action goal or outcome (Hamilton and Grafton, 2008)
5. the dorsal part (dPMC) and the ventral part (vPMC) of the premotor cortex – that contains a repertoire of premotor representations recruited both during the execution and the observation of goal-directed actions (Gallese et al., 2004) – are involved in the inference of likely action intentions (Rizzolatti & Craighero, 2004)

3. Pitfalls of the motor simulation theory

We saw that actions are hierarchically organized according to whether they can be identified or not on the basis of their motor components. Such an organization implies that the highest levels of the hierarchy – the goal and intention levels – are detached from its lowest levels: they cannot be strictly reduced to the kinematic and the motor levels. Accordingly, the strategy that consists in simulating other people’s intentions based on visual information elicited by the observed movement kinematics only, would be a suboptimal, unreliable strategy.

In a theoretical article, Jacob and Jeannerod (2005) addressed this problem by arguing that the properties of the mirror system are well designed for representing low-level, motor subgoals, such as grasping movements of an object-oriented action. On the other hand, such a system would not be well designed for representing more abstract goals and intentions, such as social or communicative intentions. To illustrate the pitfalls of the motor simulation theory, the authors presented the following thought experiment (p.23):

> “Consider Dr Jekyll and Mr Hyde. The former is a renowned surgeon who performs appendectomies on his anesthetized patients. The latter is a dangerous sadist who performs exactly the same hand movements on his non-anesthetized victims. As it turns out, Mr Hyde is Dr Jekyll. Suppose that Dr Watson witnesses both Dr Jekyll's and Mr Hyde's actions. Upon perceiving Dr Jekyll, alias Mr Hyde, execute the same motor sequence twice, whereby he grasps his scalpel and applies it to the same bodily part of two different persons, presumably the very same mirror neurons produce the same discharge in Dr Watson's brain. Dr Jekyll's motor intention is the same as Mr Hyde's. However, Dr Jekyll's social intention clearly differs from Mr Hyde's: whereas Dr Jekyll intends to improve his patient's medical condition, Mr Hyde intends

> to derive pleasure from his victim's agony. By matching them onto his own motor repertoire, an observer simulates the agent's movements. Simulating the agent's movements might allow an observer to represent the agent's motor intention. We surmise that it will not allow him to represent the agent's social intention.”

According to Jacob and Jeannerod, motor simulation would only play a minor role in the representation of the more abstract levels of actions. Motor simulation would thus be involved in representing the action’s low-level features only – the ‘motor intention’ level (Jacob & Jeannerod, 2005). In the above fictive experiment, the alternative interpretations of the observed behavior fit equally well with the observed kinematic patterns, and, more generally, with the visuomotor information conveyed by the action scene. As the current movement kinematics of either Mr Hyde or Dr Jeckyll do not sufficiently restrict the space of their candidate and mutually exclusive intentions (i.e., treating or torturing), simulating these kinematics (mapping the observed kinematics onto the observer’s motor repertoire) does not represent an optimal strategy for accurate behavioral prediction.

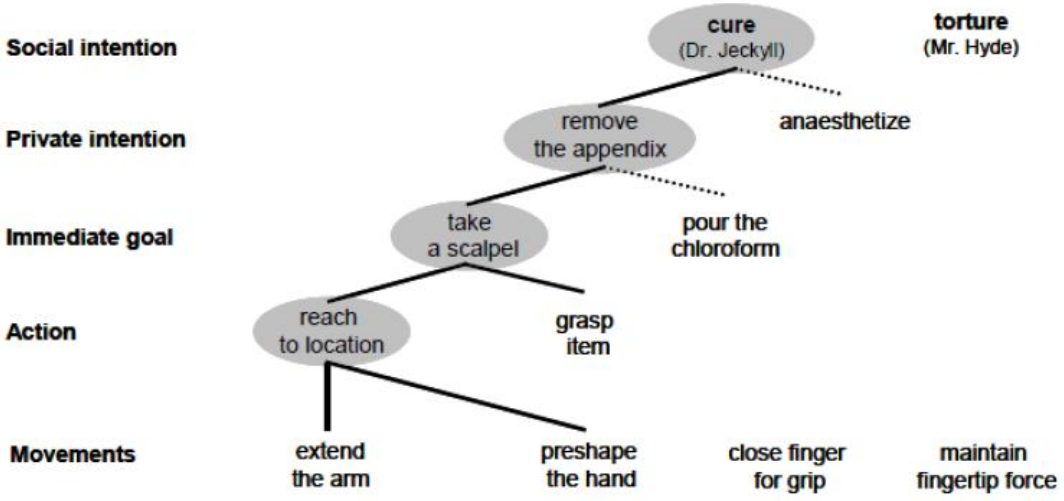


Figure 2. Figure 5. Hierarchical organisation of intentional action (inspired from Hamilton and Grafton, 2006; Chambon, 2008, with example taken from Jacob and Jeannerod, 2005). A social intention (anticipating a social interaction) could involve several private intentions (tasks goals). Each possible private intention could be composed of several immediate goals, each of which requires a sequence of basic actions. Finally, each basic actions is associated with an action that is composed of several movements. Some exemplars are shown at each level, and the dotted lines represent intermediate goals that are required to distinguish between the two social intentions ‘to cure’ and ‘to torture’.

B. Action understanding as an inverse problem: the role of prior information

As suggested by the fictive experiment designed by Jacob and Jeannerod, predicting another person's behavior is rarely a 'one-to-one' mapping problem as many competing intentions may have potentially caused the observed behavior. Conversely, one specific intention may be achieved through many different action sequences ('many-to-one' mapping problem). In both cases, predicting an intention from an observed action sequence, or predicting an action sequence from an intended goal, entails two kinds of inferences that require the observer to select the most relevant hypothesis among a set of alternatives, which best accounts for the current situation, and that can be generalized to new situations. In other words, for each observed behavior that has to be interpreted, the human observer should find an appropriate strategy that allows him to constrain the space of competing hypotheses. As it has been previously mentioned, motor simulation mechanisms might not sufficiently constrain such space.

Overall, understanding or predicting an action can thus be viewed as a kind of *inverse problem* (Kilner et al., 2007a, 2007b; Baker et al., 2009; Csibra & Gergely, 2007). Inverse problems precisely refer to situations in which the same sensory input can have many different causes, or in which a specific cause may arise from many different sensory inputs. Thus, inverse problems are ill-posed and cannot be solved by analytic methods, for the available information does not sufficiently constrain the space of candidate solutions. Inverse problems are characteristic of many situations in physics or in neuroimaging (Schmidt et al., 1999).

Inverse problem approaches have also been used to address a number of issues that the visual system has to solve, such as the perception of bi-stable or degraded stimuli, whose resolution requires appealing to prior knowledge (such as knowledge about the position and orientation of the light source) or making further assumptions about the nature of the observed stimulus (Mamassian and Goutcher 2001). Similarly, many different intentions may be achieved through the same action kinematics, and many different action kinematics may achieve the same intention. Crucially, the inverse problem addressed by the intentional inference further depends on the level of abstractness of the goal to predict or infer – i.e., on the type of mapping (one-to-one, many-to-one) between the action and the goal it achieves.

In human adults as in infants, the apparent ease of goal inference does not reflect the complexity of the processes at stake. Humans are very efficient at extracting goals from noisy, ambiguous, or even sparse sensory data (Baker et al., 2006). What are the specific mechanisms that make this inductive leap from incomplete data possible, and whereby human observers solve the inverse problem of action understanding? Just like the resolution of bi-stable stimuli by the visual system, the inverse problem of action understanding cannot be solved with the available visual input only. For example, one needs to take into account another class of information in order to select which of the competing intentions best explains the observed behavioral sequence. Such information can reflect the observer's own preferences, her/his knowledge about the social and the situational contexts in which the action takes place, her/his knowledge about the function and use of the artifact that is the target of the observed action, and most importantly, her/his specific knowledge about the observed agent's behaviors and intentions (Frith and Frith, 2006). In turn, these classes of information generate prior expectations allowing the observer to make further assumptions about the cause(s) of the observed action, or about the action sequence that may have been used to achieve the intended goal(s). By analogy, visual perception implicitly assumes that objects of the environment are illuminated from above and, if the system does not contain any extra information that enable the observer to solve ambiguities elicited by the convexity of a surface, it automatically assigns depth values to points on this surface based on this prior assumption about the position of the light source (Ramachandran, 1988). By making similar assumptions about the cause of an action, human observers may restrict the number of competing hypotheses in situations where the current sensory information under-constraints the space of candidate causes for a same phenomenon. As pointed out by Csibra and Gergely (2007), it is important to underline that assumptions used by human observers to solve inverse problems *are* assumptions, which means that they do not have to be, and not always are, valid. Crucially, in the case of an inverse problem, the validity of a solution depends on the probability that the assumption is true (Baker et al., 2006). For example, considering that natural light is generally coming from above, the assumption of 'illumination from above' is highly likely to give valid solutions. Importantly, the selection of certain assumptions also depends on the current context in which the action takes place. The 'illumination from above' assumption is irrelevant in an environment where only directionless, artificial light is present. Going back to the example of the Dr. Jekyll and Mr. Hide, the assumption favoring the 'Dr. Jekyll explanation' would instantly lose its relevance if, let's say, one heard a strident cry of pain from the victim.

In summary, understanding and predicting other people's behaviors and intentions requires the observer to make inductive inferences from intrinsically ambiguous, or sparse data – i.e., under conditions of sensory uncertainty. This uncertainty can be purely statistical – e.g.

resulting from the combinatorial explosion due to the inherent nature of the inverse problem – or more simply perceptual – e.g. resulting from a noisy or incomplete sensory signal. On the one hand, prior expectations restrict the space of candidate hypotheses. On the other hand they enable the observer to infer an action intention from a noisy situation by complementing the missing visual information (Griffith et al., 2008). Goal or intention inferences thus require prior knowledge that may be of different types and be derived from various sources.

1. Social environments

Numerous studies have shown that human observers may form prior expectations from the social intentions of their peers. For example, individuals are prone to take the social risk of helping another despite the possibility of non-reciprocation. Yet, relationships based on reciprocal altruism are inherently unstable, for an individual may be tempted to act according to short-term self-interests only, and thus, accepting the help of others without reciprocating. However, in spite of the potential cost and disavantage of engaging oneself in reciprocal interactions, humans are naturally biased to trust their relatives for they naturally expect reciprocation from other people rather than deception (Trivers, 1971). In line with this hypothesis, several studies using fMRI (Krueger et al., 2007; Behrens et al., 2008; Rilling et al., 2004) have demonstrated that engaging in reciprocal interactions - .i.e., deciding whether to trust a third party or not– involved the same brain system as the Theory of Mind network (i.e., the dorsomedial prefrontal cortex, the posterior cingulate cortex, and the temporo-parietal junction).

In the social domain, reputational knowledge also generates prior expectations that bias our appreciation of others' behavior, as well as our motivation to learn new information from them. More specifically, it has been demonstrated that the moral reputation of a social partner can outweigh direct experience in deciding whether or not to trust the partner (Delgado et al., 2005).

Finally, the perception and the understanding of other people's behavior is also modulated by other forms of social influence, such as the pressure a group may exert on an individual's decision. The modification of an individual's judgment under such pressure – i.e., an effect that is often referred to as 'social conformity' – was first demonstrated by Asch in a series of seminal experiments (1951, 1955). In these experiments, participants were asked to estimate and make judgments about the relative lengths of line segments. Although these tasks were perceptually simple, participants frequently made wrong answers when a group of peers was also giving the wrong answer. Recently, it has been shown, in a perceptual decision-making task (i.e., judging the beauty of faces) that when individual and group opinions conflict, the brain generates neural signals in the anterior cingulate cortex (a region located in the posterior medial frontal cortex) and the ventral striatum that share common features with the prediction error signal carried by the dopaminergic system, and observed during reinforcement learning (Jones et al., 2011; Campbell-Meiklejohn et al., 2010). Evidence shows that the detection of conflicting social feedback is processed by an individual as the detection of an erroneous action outcome which predicts the magnitude of subsequent behavioral adjustment (Klucharev et al., 2009). The strength of this 'prediction error' signal is correlated with the amplitude of the conflict, and is predictive of the individual's propensity to change future decisions to conform to the group (Klucharev et al., 2009). These results show that the mere presence of a group generates prior expectations that bias the observer's appreciation of the behavior of individuals belonging to this group, and alters her/his own decision criterion as well.

2. Motor expertise

Action understanding and prediction also benefits from prior knowledge acquired from past experience. Indeed, one can predict the end-state of a movement from its starting configuration, but such a prediction may also depend on the motor expertise of the observer. For example, Romani and colleagues (2003) asked basketball players and naive participants to predict the end-states of filmed basketball free throw shots. By the third frame of the video clips, basketball players were much more accurate than non-experts in predicting whether the ball would enter the basket or not. This effect of expertise has been shown to be implemented in the observer's own motor system, as revealed by an increase of the corticospinal excitability during action predictions by expert observers (Aglioti et al., 2008). This finding suggests that motor expertise helps the observer predict the end-state of an observed action (provided it belongs to their domain of expertise) by enabling a fine-grained analysis of action kinematics. Prior motor information emerges from the formation, through practice, of skillful motor programs the observer uses to anticipate the course of an action. It is of note that the two studies mentioned above do not specify the extent to which motor priors acquired from expertise bias the observer's intention predictions. Rather, these studies demonstrate that motor priors are a useful source of information for on-line monitoring of action goals whose representation is already well specified.

3. Statistical regualities of past events

Another form of prior information acquired from past experience may also play an important role in action understanding. It is well known that many of our most fundamental abilities are the result of an implicit adaptation to the statistical regularities of the environment (Reber, 1967, 1993). The ability to extract the structure of events from their frequency of occurence has been shown to generalize in a number of domain such as language acquisition or perceptual learning. Interestingly, these mechanisms of statistical learning appear to be functional at a very early stage of the cognitive development of young humans. Indeed, even infants are able to detect complex statistical patterns and adapt their future behavior accordingly (Gopnik and Wellman, 2012). For example, Saffran and co-workers (1996) investigated how 8-month-old human infants segmented words in a corpus of artificial speech. The underlying assumption of the authors was that in natural speech, adjacent sounds that co-occur with a high probability are usually found within words, whereas low probability sound pairs tend to span word boundaries. According to the authors, this difference in likelihood of co-occurrence provides potential information for word boundaries, and could contribute to early language acquisition by strengthening the ability to segment the speech into meaningful units.They showed that young infants were able to extract statistical patterns in auditory input that are based on transitional probabilities defining the sequencing of the input's components. Similar results were also obtained with the learning of sequences of discrete visual stimuli whose ordering followed a statistically predictable pattern (Fiser and Aslin, 2002a,b), and so even with younger children of 2 months of age (Kirkham et al., 2002). Furthermore, a recent study by Wu and collaborators (2011) showed that once the statistical patterns have been extracted from the co-occurence of visual stimuli, 9-months-old infants are capable of forming expectations about visual objects and use them to make inference about their properties (see also Kersten et al., 2004).

The variety of cognitive domain in which statistical learning appears to be efficient suggests that these learning strategy is ‘domain general’, that is to say, can potentially be extended to the learning of many other forms of stimuli (Kirkham et al., 2002). As such, there is no reason that action understanding and action prediction would not benefit from statistical information derived from past experience of the observer (Tenenbaum et al., 2011). Indeed, accumulating information about the number of times a specific action sequence is followed by a specific goal, or about how frequently an action sequence has been performed to achieve a particular intention, may actively shape predictions made by an observer.

C. Solving the inverse problem: rationality, simple heuristics and probabilistic inference

All these sources of prior information – i.e., social environments, motor expertise, and statistical structure of past events – can be used by an observer to constrain the space of candidate intentions that potentially drive the behavior of the observed agent.

Over the last decade, a growing number of studies have investigated the type of *inferential* mechanisms involved in action understanding and prediction. In the next paragraphs I will outline two types of inferential mechanisms, and I will distinguish them on the basis of both their computational complexity and the type of information computed. These two classes of mechanisms may interact in action understanding and action prediction. The first type of mechanism can be termed ‘heuristic’ by analogy with Artificial Intelligence. Heuristics provide the advantage of being computationally tractable, though restricted to a narrow domain (Chase et al., 1996). We specifically aim to describe a heuristic that plays an important role in action understanding: the naive theory of rational action (Gergely and Csibra, 2003) whereby observers assign goals to agents through estimating the *optimality* of their actions. The second type of mechanism is the Bayesian probabilistic inference, which provides the advantage of being very flexible, though computationally more complex (Baker et al., 2006; Baker et al., 2009). Interestingly, these two types of inferential mechanisms – i.e., Bayesian inference and simple heuristics – share a similar basis: both should be rational in some way, and both are guided by a rational interpretation of the observed action (Gergely and Csibra, 2003; Baker et al., 2006, 2009).

1. About rationality

It is important to delimit the notion of ‘rationality’ prior to going further. Generally, the term ‘rationality’ defines a particular causal relationship that links the components of reasoning and action. Specifically, it refers to the idea that the means used to conduct reasoning or to perform an action are appropriate given the desires that motivate them. Thus, the notion of ‘rationality’ crucially depends on the notion of ‘goal-directedness’. Of note is that rationality refers to a very specific type of causal relationship that one can distinguish from the causal relationship through which physical events are usually described. For example, there is no reason to believe that a planet behaves rationally when it follows its orbit. Here, the causes that describe the relationships between the components of such a physical system are formal and efficient, but not ‘finalistic’. It has long been argued that behaving rationally – i.e., choosing the appropriate means to optimize the chance of achieving a desired goal – amounts to applying the rules of the probability theory (Laplace, 1814). In the next paragraph I will explain that such approach is, however, no longer adapted to explain human action planning, execution, and prediction. Rationality of human behavior and decision-making is not absolute but limited, bounded. In other words, rationality is contingent and depends on the internal constraints of agents, as well as on external constraints of the environment.

1.1. Rationality is bounded

Humans and other animals make inductive inference. Firefighters predict how fire will progress from various cues such as smoke and roof ‘sponginess’, while peahens rely on how elaborated peacocks' tails are to infer their fitness before deciding whether to mate or not (Petrie, 1994). For firefighters and peahens alike, the cues which their inferences are based on are typically uncertain. So how can organisms make rational inferences based on uncertain cues? According to Peterson and Beach (1967),

> "Probability theory and statistics can be used as the basis for psychological models that integrate and account for human performance in wide range of inferential tasks." (p.29)

Following this *classical* view, human reasoning can be evaluated within the norms provided by probability theory: if the outcome of reasoning diverges from norms of the probability theory, one may conclude that there is something wrong with such reasoning, not with the norms. If there is a mismatch between the two, then the assumption of rationality is broken.

Suppose you and four of your friends buy lottery tickets consisting of six number combinations each. Suppose that the combination of numbers is randomly assigned to each ticket. Suppose that, on the five tickets you and your friends bought, only one is composed of successive numbers – let's say '22-23-24-25-26' –, and the four others are composed of non-successive distributed numbers – let's say '3-9-14-20-33-42'. In this situation, the probability theory clearly argues that each possible combination is as likely to be a winning combination as another, due to the law of random draws independence. Thus, a combination composed of successive numbers has the same chance of winnig as a combination involving non-successive numbers. Now suppose you ask each of your friends to choose a ticket among the five you bought. The ticket that will remain to you will depend on the choices of your friends. Under such circumstances, it is highly likely that you ‘inherit’ a ticket with six successive numbers – being likely that your friends had chosen tickets with *non*-successive numbers. Indeed, agents behave in this situation as if they believed that the better distributed combinations would have more chance to win. Even though people know the law of random draws independence (i.e., each number combination has an equal chance to win), the belief that choosing a non-successive number distribution is more advantageous persists. This example shows that many of our daily inferences do not follow the rules of probability theory, especially in situations where the weighting of every possible alternative leads to a combinatorial explosion. As such, “rationality” rules do not even approximate human behaviors and inferences, for the human brain has not evolved to perfectly understand the rules of probability theory. In such circumstances, no wonder some researchers interpreted human behaviors and inferences as non-rational, for the human brain is not a super computer adapted to embody the rules of probability theory. For example, Slovic and co-workers wrote (Slovic et al., 1976):

> “It appears that people lack the correct programs for many important judgmental tasks.... it may be argued that we have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.” (p.170)

Yet, given the success of the human interpretative skills one has to admit that, to a large extent, humans are ‘rational’ in some way. The example of lottery tickets suggests that humans have a strong tendency to attribute decision-relevant properties to objects that, in fact, do not have any. This tendency highlights the need of finding cues rendering our choices satisfactory, guided by a specific ‘reason’ or, so to say, rationale. In this context, the example of lottery tickets shows that our inferences are *biased* by our beliefs and other prior expectations about events, whether they be valid or not.

According to some authors (Chase et al., 1996; Newell, 2005; Simon, 1990), the classical view of rationality neglects this crucial aspect of behaviors as it assumes that rational inference is blind to our expectations regarding the content of the processed information and the context of the task. The alternative view on rationality posits that, given the computational limitedness of the human mind, human rational inference is necessarily bounded by a number of constraints associated with the agent itself or with the context of the task (Gigerenzer and Gaissmaier, 2010; Griffiths et al., 2010).

Rationality is thus limited and bounded (Simon, 1990) in the sense that the agent's behavior is strongly constrained by cognitive capacities, time, and structure of the environment such as the search space of a task (i.e., the number of alternatives that can be used as relevant solutions to solve a task problem; see Acerbi et al., 2011, 2012). A decision can thus be interpreted as rational when it is *satisfying* enough, given the state of these constraints.

Inference-making mechanisms also depend on the nature of these constraints. For example, imagine someone who has to infer the emotional state of a person by scrutinizing her facial expressions – let's say a smile. What kind of computation is her brain performing? If the information she has is limited to only the perception of the face, without any additional knowledge, her brain will use a simple recognition heuristic for the detection of facial expressions, and she will conclude that the observed individual is happy. Here, the heuristic underlying the inference consists of a simple cognitive module that generates prior expectations filtering the sensory inputs. If the inputs pass the filter, prior expectations are satisfied, and her brain considers the face as expressing happiness (Sperber and Hirschfeld, 2004). This kind of recognition heuristic provides a rapid and parsimonious solution to the task problem, and is likely to be valid in a very large number of situations (possibly providing arguments for why such a module may have evolved through natural selection). Now suppose that the very same person usually *feigns* happiness to practice deception, but the observer is not aware of that. Her brain will use the same (invalid) recognition heuristic. Finally, suppose the observer and the observed individual are colleagues: the former knows from experience that the latter usually feigns happiness to practice deception, especially in the case where a personal interest is at stake. In this context, the mechanism producing the inference will be more complex and flexible, so that expectations generated by the face recognition heuristic are weighted by prior expectations that have been formed through probabilistic sampling of past observations, and by the context of the situation. In any of these cases, the decision that is made about the psychological state of the other person is rational, whether it is based on simple heuristics or more complex probabilistic inferences. Such rationality is determined by a set of internal and external constraints – such as the cognitive capacity of the decision-maker, the knowledge she possesses about the observed agent, and the environmental context in which the situation takes place.

1.2. The principle of rationality

Strikingly, it seems that the computational principles of bounded rational inferences are the same as those driving our assumptions about the rationality of an observed action. According to Dan Dennett (1971; 1991) humans do not interpret other people's behavior as any other physical system, but as an ‘intentional system’ endowed with specific desires, beliefs, intentions, etc. As such, our interpretation of a person’s intention is derived from an estimation of the rationality that characterizes her/his action. Such estimation is not guided by the rules of probability but mostly determined by situational constraints and by intrinsic characteristics of the agent. Put another way, humans assume *by default* that other agents behave rationally given situational constraints: this is what Dennett called the ‘principle of rationality’. According to him, this associated strategy – termed ‘intentional stance’ – provides a powerful predictive model of living organisms’ behavior. This model would be much better than, for example, a predictive model based on physical causality.

To illustrate his idea, Dennett (1987) imagines a martian who observes the final of the american football championship – the Super Bowl. Suppose that an inhabitant of Mars points a very powerful telescope at Earth and, by chance, falls on the field where the Super Bowl is currently being played. Suppose that our martian is omniscient in perceiving and predicting events following a physical causal scheme. Our martian observes human-shaped entities running to and from the field, throwing a ball, chasing after each others, violently punching each others, etc. Dennett argues that if our extraterrestrial observer cannot apply the scheme of ‘intentional pattern’ within the principle of rationality, he will not be able to perceive the observed events as stages of a game, that is to say, as stages of a joint, intentional activity in which events have a particular ‘reason’ to occur. If our martian does not discover this interpretative scheme, he will only appreciate the physical aspects of events: movements, trajectories, collisions, etc. But let’s suppose that he manages to adopt the “intentional stance” so that he is now able to enjoy the show for what it is: he will understand why one of the human-shaped entities jumps up and down after having crushed the ball behind a seemingly arbitrary line, he will further understand why only the cheerleaders of one side burst out and those of the opposite side do not, etc.

According to Dennett the human brain has been tuned by natural selection to perceive ‘intentional patterns’ in the behavior of humans and other animals. Dennett further remarks that some humans also use the intentional stance to attribute rational causes to non-biological events, such as the hand of God in explaining the occurrence of natural catastrophes. Recently, research in the field of cognitive neuroscience identified in young human infants a very similar interpretative strategy (Csibra et al., 1995). These data reveal a very early aptitude of humans to infer goal-directed, intentional patterns from biological and non-biological events. This heuristic would assume *a priori* that observed behaviors are rational, given the constraints of the situation and the internal properties of the agent.

2. A simple heuristic for goal attribution: the naive theory of rational action

A number of studies have demonstrated that humans are equipped with an interpretative skill that enables the perception of observed actions as goal-directed, independently of whether actions involve, or do not involve, biological movements (Gergely and Csibra, 2003). Strikingly, it has been shown that this skill operates very early during the development, around the end of the first year of birth (Gergely et al., 1995).

For example, Gergely and colleagues (1995) habituated twelve-month-old infants to a computer-animated goal-directed event during which a small circle approached and reached a large circle by jumping over (‘means act’) an obstacle between them (‘situational constraint’). During the test phase, experimenters changed the situational constraint by removing the obstacle. Infants were then presented two test conditions: one in which the small circle approached and (contacted) the large circle by jumping as in the habituation phase (but without the presence of an obstacle), and a new one in which the small circle went straight towards the large circle. Researchers found that infants looked longer at the jumping action, but showed no dishabituation to the new straight action. The increased looking time registered during the jumping action indicated a violation of infants' expectations. This violation of expectations was due to the fact that the small circle used a suboptimal, inefficient means to achieve its goal (contacting the large circle) as there was no obstacle to jump over. In contrast, the fact the the looking time did not increase during the straight action was interpreted as a satisfaction of infants' normal expectations: the straight action appeared as the most efficient means to achieve the goal in this situation. These results are interesting for several reasons. First, they reveal that very young infants are able to perceive actions as goal-directed. Second, they are able to evaluate which alternative, according to the constraints of the situation, provides the most efficient means to achieve a goal. Finally, infants expect the 'agent' to adopt the most efficient means available to achieve a given goal.

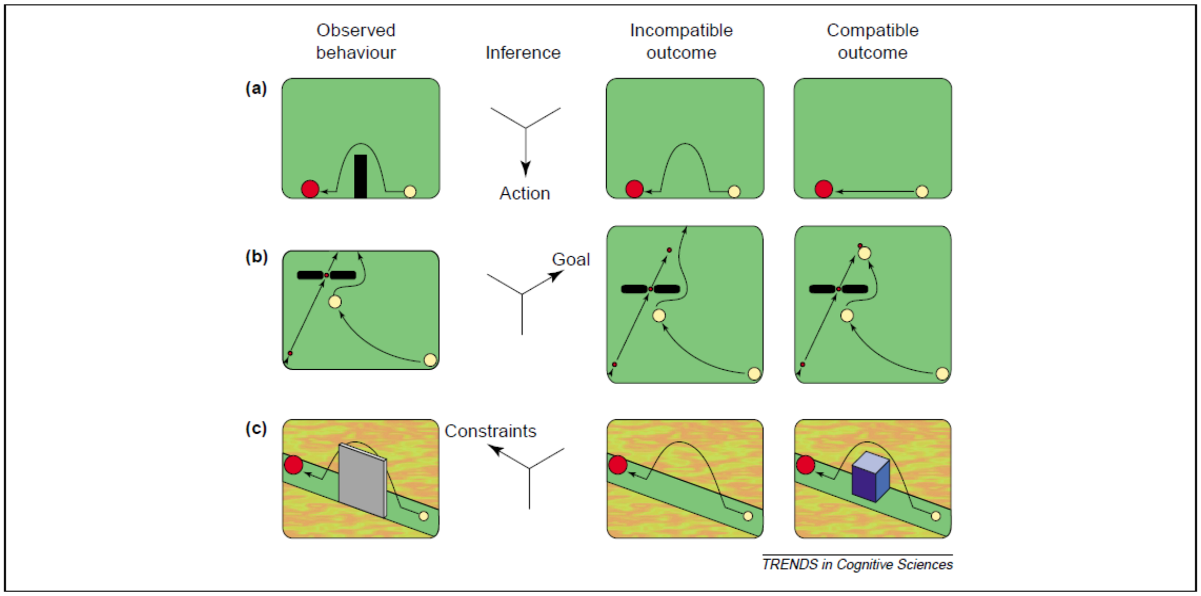


Figure 3. Representation of the three types of inference that infants can make on the basis of a teleological representation of actions (taken from Csibra and Gergely, 2003). One-year old infants were habituated to the event depicted in the left column (Observed behaviour). Their interpretation of this event was tested by presenting them with two different outcomes, one of them being incompatible (the middle column), the other one being compatible (the right column) with a possible inference based on a teleological representation of the event. Infants looked longer at the incompatible outcome than the compatible outcome events, indicating that they based their inference on a teleological interpretation of the observed behaviour. The inference of action means (a) has been demonstrated by Gergely et al. (1995) and Cisbra et al. (1999). The inference of action goals (b) and situational constraints (c) have been demonstrated by Csibra et al. (2003).

According to the authors, this early sensitivity to action goals and means cannot be explained by either the framework of the ‘theory-theory’ or by the ‘simulation theory’ alone. Instead, Gergely and Csibra propose that young human infants can “represent, explain and predict goal-directed actions by applying a non-mentalistic, reality-based action interpretational system, the ‘teleological stance’ ” (p.289). The teleological stance is a primitive version of the intentional stance proposed by Dennett. Like the intentional stance, this interpretative system is hard-wired, and drives the inference of teleological (i.e., goal-directed) – rather than causal – relationships between action sequences, future goals, and current situational constraints. The advantage of this system is twofold. First, it enables young individuals to interpret observed actions as goal-directed without using a theory about how other people minds work (through reference to propositional concepts such as ‘belief’, ‘desire’, or ‘intention’). Second, it enables young individuals to interpret actions of non-biological agents (e.g., a circle animated on a computer screen) as goal-directed, without any reference to motor simulation or resonance mechanisms. Furthermore, the authors argue that neither the theory-theory nor the simulation theory provide sufficient arguments to account for the type of mechanisms whereby young infants detect the relevant aspects of a current situation and use these aspects as the basis for their interpretation of the observed action. The mechanism would simply rely on a simple but central axiom: the rationality principle.

As described above, the rationality principle provides a powerful predictive model of events. This model allows human agents to infer a specific type of causal relationship that other predictive strategies (e.g., strategies based on physical cues) do not permit. Following this model, human observers expect others' actions to be the most efficient means of achieving a desired goal, given the current constraint of the situation. An action can be considered as “efficient” when it minimizes the cost of motor parameters, for example, or when it optimizes the probability of properly achieving the desired goal. The rationality principle is an axiom that relates to each of the other three types of intentional mental states that may be attributed to an agent: her beliefs (‘she believes that’), her desires (’she wants to’), her intentions (‘she intends to’). The content of these three types of mental states represents the three main aspects the teleological stance relies on: the situational constraints (‘she believes that’), the goal (‘she wants to’), the action (‘she intends to do’). One can thus imagine an organism without any naive psychology that would benefit from a teleological interpretative mechanism. This organism would be able to evaluate an observed action as the most efficient mean to achieve a goal simply by applying the rationality principle to the relevant aspects of the current situation.

The rationality principle thus serves two functions. First, it generates prior expectations about how “appropriate” an observed behavior is. Second, these prior expectations can be used to restrict the space of possible causes for an observed action (see figure 3): given the information available about the specific contents represented by any two of the three aspects of the reality (e.g., goal and situational constraints), one can infer what the content represented by the third aspects ought to be (e.g., action means) (Gergely et al., 1995; Csibra et al., 1999; Csibra, 2003).

In summary, human infants are equipped with a naive theory of rational action, similar and the one postulated by Dennett. This theory drives the interpretation and the prediction of other people’s actions. According to Gergely, Csibra and co-workers, the content of this naive theory is essentially non-mentalistic, and accounts for goal attribution without motor simulation. It is based on the rationality principle, which enables the guidance of everyday inference about action goals, means, and situational constraints in which an action takes place. This ‘teleological stance’ would be the pre-requisite for using a more complex, mentalistic interpretation of actions. Recent evidence suggests that this interpretative strategy is applied in a variety of contexts (Csibra, 2003, Wagner and Carey, 2005) and convincingly accounts for the behavior of different sorts of ‘agents’: human-like (Sodian et al., 2004; Kamewari et al., 2005) or not (Luo and Baillargeon, 2005; Bíró and Leslie, 2007); biologically plausible or not (Southgate et al., 2008). Remarkably, it has also been shown that ‘rationality’ of an action was also used by human infants as a key characteristic to decide which behaviors to imitate or not (Gergely et al., 2002).

Together, these data show that interpreting other people's behavior is not primarily driven by motor properties of the underlying action. Rather, prior expectations, such those elicited by the teleological stance, play a critical role in our ability to understand and predict other people's behavior. Crucially, this ability can be achieved through very simple heuristics.

3. A flexible mechanism for goal attribution: Bayesian probabilistic inference

In the last past few years, some researchers investigated the apparent ease of goal inference made by human infants and adults. In particular, Baker and co-workers (2009) have consistently argued that, given the inverse problem of goal attribution (i.e., the fact that the same sensory input may have many different causes), goal inference must involve complex computations, rather than simple heuristics.

Furthermore, according to these authors, accounting for the problem of action prediction with classical statistical learning mechanisms – i.e., recording over the infinite number of experienced action situations all the ways any number of intentions can be achieved and all the ways they effectively are achieved – would be almost impossible. Instead, action understanding and action prediction require a more flexible approach.

Starting from the assumption that everyday inferences about other people's intentions are made in noisy or uncertain situations (due to the hidden nature of intentions and to the intrinsic scarcity of perceptual inputs), Baker and collaborators claim that such inferences can be convincingly described according to the principle of Bayesian probabilistic inference. Griffiths et al. (2010) describe the basics of Bayesian probabilistic inference as follows:

> “if a learner considers a set of hypotheses H that might explain observed data d, and assigns each hypothesis $h \in H$ a probability $p(h)$ before observing d (known as the ‘*prior probability*’), then Bayes’ rule indicates that the probability $p(h|d)$ assigned to h after seeing d (known as the ‘*posterior probability*’) should be:
>
> $$\boldsymbol{p(h|d) = \frac{p(d|h)\, p(h)}{\sum_{h\, \in\, H} p(d|h)\, p(h)}}$$
>
> where $p(d|h)$ is the ‘*likelihood*’, indicating the probability of observing d if h were true, and the sum in the denominator simply ensures that the posterior probabilities sum to one. Bayes’ rule thus indicates that the conclusions reached by the learner will be determined by how well hypotheses cohere with prior knowledge, and how well they explain the data.” (p.358)

Thus, Bayesian probabilistic models highlight the role of prior knowledge in accounting for how people learn from noisy, or incomplete data, and provide a framework for explaining precisely how prior knowledge interacts with data to guide inductive inferences. Remarkably, these models do not require prior knowledge to be *innate*, such as in the case of knowledge derived from our intuitive theories (Clark, in press). Prior knowledge can also be learnt from observed past experience through extracting the probablistic structure of events. Following this line, Baker et al. (2009) demonstrated in a series of experiments that solving the inverse problem of action understanding precisely requires prior knowledge about the structure and the content of agents’ behaviors and intentions. Specifically, it requires the ability to navigate through a large space of possible interpretations and to infer the best candidate alternative.

Put in the context of predicting an agent’s intention from perceived movement kinematics, the Bayesian probabilistic inference combines two parameters: the prior probability [$p(h)$] – i.e., the probability of observing the achievement of a possible intention (the set of hypotheses H) before the occurrence of the next behavior (the observed data d); and a perceptual information having a certain likelihood [$p(d|h)$] – i.e., the likelihood of perceiving a certain behavioral sequence (d) given that a certain intention (h) is true. From the combination of these two parameters follows a third parameter, which is the outcome of the Bayesian inference: the posterior probability $p(h|d)$ – i.e., the probability that the observer's assumption is true (h) given the currently observed behavior (d). Thus, predicting the achievement of a certain intention would consist of combining prior knowledge about the agent's intentions (acquired from probabilistic sampling of past events) with the perceptual evidence accumulated over time.

This is precisely what we (Chambon et al., 2011b) showed in a recent study (see the published version of the article in appendix n°1). This study aimed at investigating action prediction in healthy adults and in patients suffering from schizophrenia, who are known to be impaired at understanding the intentions of other agents. The experimental setting consisted of four action prediction tasks in which participants observed a filmed agent performing several types of actions which varyied according to the abstractness of the achieved goal as well as to the target of the action. In the first two tasks participants were required to infer the basic (i.e., the simple goal of a motor act) and superordinate intentions (i.e., the general goal of a sequence of motor acts) of the observed agent. In the third and the fourth tasks, they were required to infer the agent's *social* basic and *social* superordinate intentions (i.e., simple or general goals achieved within the context of a reciprocal interaction). In each of these tasks, both prior expectations about the observed agent's intentions and perceptual information were systematically varied. Prior expectations were manipulated by varying the probability of observing the agent achieving a certain intention, at the expense of other competing ones. Perceptual information (the movement kinematics of the observed agent) was manipulated by varying the completeness of the action sequence. The underlying hypothesis was that intention inference of healthy individuals depends on a consistent interaction between visual information and prior expectations. Accordingly, intention inference of schizophrenic patients was expected to reflect an abnormal interaction between prior information and perceptual evidence.

As expected, intention inference in healthy individuals reflected an interaction between prior expectations and perceptual evidence. The inferences were indeed contingent upon an adaptive interplay between these two sources of information, with healthy participants tending to progressively rely more on their prior expectations as the reliability of perceptual evidence decreased, and vice versa. Crucially, this interaction varied according to the target of the intention to be inferred, with prior information gaining priority over perceptual evidence when inferring intentions within a social context rather than a non-social context. In patients with schizophrenia, results showed no impairment at predicting non-social basic intentions, but they were impaired for more abstract types of intentions – i.e., non-social superordinate intention, and both types of social intentions. As expected, these impairments were associated with abnormal interactions between prior information and perceptual evidence. In the non-social superordinate condition, schizophrenic patients massively relied on their prior expectations, whilst disregarding sensory evidence. In contrast, social conditions prompted exactly the opposite pattern, with patients exhibiting weaker dependence on prior expectations whilst relying strongly on perceptual evidence.

This study first showed that prior information acquired from probabilistic exposure and perceptual information conveyed by movement kinematics interact to drive action understanding. Second, we found that the shape of this interaction was modulated by the type of intention to infer. The greater the action goal was delayed (superordinate intentions), and the more abstract it was (social intentions), the less perceptual information conveyed by the observed kinematics was prone to reduce the space of competing intentions. Accordingly, the more motor possibilities there are to achieve an intention (many-to-one mapping), the more observers relied on their prior expectations to infer it.

Interestingly, these results may help reconcile the two major accounts of action understanding developed over the last two decades, namely, the simulation theory and the theory-theory. According to the simulation theory, understanding other people's intentions requires simulating the observed action via the activation of our own motor planning system. The result of this simulation process is the selection, in the observer's own motor repertoire, of the intention that may have caused the very same action. This explanation highlights the role of visuomotor information extracted from the observed kinematics. According to the theory-theory, on the other hand, action understanding is based on an inferential mechanism that emphasizes the contribution of context-related prior knowledge derived from our intuitive theories of human behavior (e.g., the naive theory of rational action proposed by Gergely and Csibra, 2003) as well as on the observer's past experiences and rules she/he has drawn from them. While apparently opposed, the data from Chambon et al (2011) are consistent with several other recent studies which together plead in favor of a complementary role for simulation and inferential mechanisms in action understanding (Brass et al., 2007; Keysers & Gazzola, 2007; Kilner et al., 2011; de Lange et al., 2008). Taken together, these studies suggest that intention attribution may rely on an adaptive balance between bottom-up sensory and top-down prior information, whose equilibrium would be modulated by the type of intentions to be inferred, depending on their temporal and structural complexity – from basic (one-to-one action/intention mapping) to superordinate (many-to-one action/intention mapping) and social intentions.

D. Understanding tool-use actions: how object affordances help solve the inverse problem

The simulation and Theory-theory are expected to account for the understanding of a variety of actions. Yet, it is of note that a majority of studies that investigates motor simulation mechanisms uses a specific class of complex actions as stimuli. This class of complex actions concerns tool-use behaviors (see for reviews, Rizzolatti and Craighero, 2004; Rizzolatti and Sinigaglia, 2010). Tool-use refers to a type of behavior that consists in manipulating "external objects with the goal of altering the physical properties of another object, substance, surface, or medium, via a mechanical interaction", or that consists in "mediating the flow of information between the tool user and the environment" (St Amant and Horton, 2008, pp.1203).

Tool-use behaviors occupy a particular place in human and animal behavior and deserve a particular attention for several reasons. First, the modern human environment is overrun by technology, to such an extent that a majority of our daily actions are directed towards objects, tools and other artifacts. Second, tool-use behaviors are the most frequently observed markers of cultural traditions, both in human societies and in the wild (Whiten et al., 1999; Whiten et al., 2005). Third, tool-use behaviors is often take as a comparative model to assess the ability of individuals to understand and learn socially from their conspecifics (Whiten, 2011). Finally, the mastering and manufacture of tool-use is frequently summoned to draw hypotheses about human cultural and cognitive evolution (Sterelny, 2003a; Stout and Chaminade, 2012).

1. Attributing a function to a tool requires mechanisms of intention inference

Tool-use refers to a type of behavior that "consists in manipulating external objects with the goal of altering the physical properties of another object, substance, surface, or medium, via a mechanical interaction, or that consists in mediating the flow of information between the tool user and the environment" (St-Amant and Horton, 2008). Understanding tool-use behaviors requires the acquisition of knowledge about its use and function from observing action means and goals achieved by tool users. As such, tool-use behaviors cannot be understood at the level of motor intentions only, where simple object manipulations are involved. Instead, the achievement of a tool's function refers to a particular instance of "super-ordinate" intention: the object itself is not the target of the intention but just a proxy to achieve a temporally distant goal. As such, tool-use behaviors are a particular incarnation of the inverse problem. They cannot be solved through merely analyzing the motor sequence underlying the observed behavior. A functional intention can be achieved using many different motor sequences, and a particular motor sequence can be performed with the aim of achieving many different functional outcomes.

As Csibra and Gergely observed (2007), attributing a function to a tool amounts to 'sticking' a particular action goal to this tool. As such, the function of an artifact is a 'frozen' goal (a tool can serve a similar goal, or function, under a large scope of situations). Obviously, understanding and predicting tool-use behaviors from observation would thus engage the very same mechanisms as those described in the preceding sections, namely, i) a mechanism of visuo-motor transformation subserving the simulation of the observed tool-use action, ii) a simple heuristic enabling goal attribution with regard to the rationality of the observed tool-use action, and, finally, iii) a more elaborate inferential mechanism based upon prior knowledge acquired from probabilistic sampling of past observations. Together, these mechanisms would enable human observers to derive knowledge about the possible uses and functions of a tool from observing goal-directed, intentional movements performed by an agent.

2. Attributing a function to a tool requires the detection of object affordances

In addition to these mechanisms, human observers may benefit from another type of heuristic to understand and predict tool-use behaviors. Such an heuristic is based upon the detection of low-level, local sources of information, such as the manipulative properties of objects, namely, their '*affordances*' (Gibson, 1979). Affordances are not intrinsic properties of objects. Rather, an affordance defines a relational property that emerges from matching the perceived physical features of an object (e.g. size, shape, texture, density) and the agent's sensorimotor constraints, her goals, plans, values, beliefs, and past experiences (Norman, 1988). Specifically, affordances 'suggest' how an agent might interact with an object, given the sensorimotor constraints of the acting organism. According to Gibson (1979), they represent the basis which any living mobile organism relies upon to guide its actions in the world. For example, the vision of a wooden stick planted in the ground, of its size and shape, may prompt the action of grasping, whereas its density and texture make it perfect to take a piece of fruit down from a tree. For birds, however, the very same stick may afford the action to rest on its upper extremity, avoiding the threat of terrestrial predators.

To underline both the similarities and the differences from the affordance notion elaborated by Gibson, Ellis & Tucker (2000) proposed the notion of ‘micro-affordances’ to refer to the potentiated elements of an action. Microaffordances are brain assemblies that represent objects; they are the product of the conjoining, in the brain, of visual responses and action - related responses that have developed throughout individual and species history as part of the process of adapting to the environment. Similarly to the original notion of affordances, microaffordances are elicited automatically, independent of the goal of the actor. Thus, microaffordances do not pertain to complex actions, which are probably mediated by the actor’s goal, such as drinking. Rather, they facilitate simple and specific kinds of interaction with objects. These simple interactions with objects also imply the activation of conceptual knowledge. In fact, microaffordances differ from Gibsonian affordances in that they are much more specific and are a consequence of object-based attention (Vainio et al., 2007). They do not elicit grasping, but a specific component of grasping, which is suitable to a particular object. For example, a ball is represented by making accessible the information that it can be reached and grasped, in order to play with it.

Some recent findings suggest that the detection of affordances is implemented within motor simulation mechanisms similar to those that are involved in action observation and action understanding. Specifically, a population of neurons mostly located in the anterior intra-parietal lobule (AIP) and F5 area of the monkey brain – the canonical neurons – has been shown to be specifically recruited when individuals deal with geometrical objects. By testing responses of single neurons located in area F5 of the monkey brain during object-oriented motor actions, Rizzolatti (1988) and colleagues have shown that F5 neurons selectively respond to different types of hand prehension movements (e.g., precision grip prehension, whole-hand prehension). Canonical neurons in areas F5 and AIP have also been shown to selectively code for visual properties of three-dimensional objects (size, shape, and orientation) in a corresponding motor format (Murata et al., 1997; Murata et al., 2000). In support of these findings, inactivation studies showed that artificial lesions of the monkey's area AIP (Gallese et al., 1994) or F5 (Fogassi et al., 2001) elicited a substantial deficit in the ability to move the hand congruently with the object size and shape. Binkofski and coworkers (1998, 1999) showed in human subjects that lesions induced in the anterior part of the lateral bank of the intraparietal sulcus led to deficits of hand shaping for grasping objects similar to those observed in inactivation studies with monkeys. The presence of canonical neurons in the intraparietal and ventral limbs of the precentral sulcus of the human brain has also been suggested (Grèzes et al., 2003). The activity of canonical neurons is thought to generate a set of motor primitives that are recruited in any type of motor action involving effector/object interactions (Flash and Hochner, 2005).

These motor primitives would participate in the generation of a set of structured motor patterns which affordances might be primarily elaborated upon. However, something more is needed to make affordances *predictive* cues of the object's function.

3. Affordances, sensorimotor constraints, and the principle of rationality: the emergence of biomechanical priors

The role of affordances in the individual exploration of objects and the discovery of their function is well known (Osiurak et al., 2010; Visalberghi et al., 2009; Whiten et al., 2004). In a recent theoretical commentary (Jacquet et al., 2012a, in press), we proposed that the primary function of affordances is just like other types of prior information - to narrow the space of possible motor acts that one can perform on a given object or tool.

Our proposal is that affordances generate effector-dependent, biomechanical prior expectations which are in line with the agent's sensori-motor constraints. Crucially, this type of prior information would emerge through a primitive form of the rationality principle that would bias agents to act towards objects with the aim of biomechanical optimization. Interestingly, it has been shown that both human and non-human primates indeed favor object-directed behaviors that minimize the muscular and/or articulator costs, given the object's affordances and the desired outcome (Rosenbaum et al., 1992; Rosenbaum et al., 1996; Sartori et al., 2011; Weiss et al., 2007). Biomechanical priors elicited by object affordances would thus provide a simple heuristic for inferring the function and use of a tool. For example, based on the amplitude of the observed agent's grip aperture and the orientation of her wrist, as well as on the size, the shape, and the texture of the object to be grasped, one may predict whether this object is meant to be lifted, pushed, or merely transported (Chambon et al., 2011). Biomechanical priors elicited by affordances may exert strong influences on inference-making, as the observed agent is expected to adopt a behavior that minimizes biomechanical costs. Therefore, the understanding and the prediction of tool-use actions should be facilitated when the observed agent uses a tool in a way that fits the observer's biomechanical expectations (low cost behaviors), and should be jeopardized in the case where these expectations are patently violated (high cost behaviors). In summary, understanding and predicting tool-use behaviors performed by a third party recruit several interacting mechanisms. First, a mechanism of motor simulation that, articulated around the axiom of rationality, enables observers to estimate the biomechanical costs of an action. This estimation would then bias the selection of the action that, among competing alternatives, is optimal given the object configuration and the desired goal. Second, a mechanism of probabilistic inference enables the observers to derive knowledge from past observations. Such prior knowledge would participate in the specification of its function and use. These mechanisms may combine when prior knowledge and biomechanical priors elicited by object affordances converge. However, they may also compete when these classes of prior information actively *diverge*; that is, when the agent's behavior violates the biomechanical expectations of the observer whilst being congruent with past observations – and vice versa.

Figure 4. Like humans, non-human animals can take advantages of object affordances to guide their actions in the world. Affordances provide cues that narrow the space of possible motor behaviors an individual can execute, given its biomechanical architecture, making possible the achievement of complex behaviors, such as tool-use, at low cost. The left and the middle photographs are taken from Seed and Byrne, 2010; the right photopragh is taken from Weiss et al., 2007.

E. Hybrid model of action understanding

Throughout the present theoretical introduction we have examined several types of mechanisms that could help in solving the inverse problem associated with action understanding, and have isolated two types of inference mechanism, each being dependent on the assumption of rationality and varying according to their computational complexity. In particular we have discussed how simple heuristics – teleological stance and affordances –, as well as probabilistic inference, provide the observer with an information that movement kinematics alone cannot provide. Thus, the teleological stance enables one to derive prior expectations about the well-formedness of an observed behavior. Affordance-based heuristics rely on biomechanical prior expectations that emerge from matching the physical properties of an object with sensorimotor constraints shared by both the observer and the observed agent. These biomechanical priors over-weight action goals (or tool functions) whose achievement minimizes biomechanical costs. Finally, Bayesian probabilistic inference allows an observer to predict an agent's intention by combining prior knowledge about the agent's intentions (acquired from statistical regularities of past observations) with perceptual evidence conveyed by the action scene.

These three types of inference contribute to solving the inverse problem of action understanding by constraining the space of possible intentions that may have caused the observed action. According to the type of action being predicted, these three respective mechanisms are informed by either sensory (movement kinematics and situational constraints) or more abstract cues (goal representation), and even propositional information (the degree of belief in one particular intention). These mechanisms are somehow compatible with both a simulationist and a theory-theory account of action understanding, emphasizing the relevance of a hybrid model that would clarify, at both cognitive and cortical levels, their mutual influence during inference-making.

1. The 'predictive coding' model

The latest version of this 'hybrid' model has been recently proposed by Kilner and collaborators (2007a; 2007b; 2011). The "predictive coding" model accounts for the adaptive interaction of two sources of information that are involved in action understanding: the prior information (being acquired from past experience, or being 'innate') and the visuo-motor information conveyed by the action kinematics. This interaction is explained within a hierarchical model of action understanding (Grafton and Hamilton, 2007). Note that the formalism used by the model – the predictive coding – is particularly adapted to account for how an observer solves the inverse problem of action understanding (e.g., many intentions can explain a particular motor sequence, or various motor sequences can be peformed to achieve the same intention).

Kilner and colleagues first point out the lack of empirical evidence for the role played by mirror neurons (and, by extension, the Action Observation Network – also termed 'mirror system') in the *human* ability to understand other people's behaviors, which has led some researchers to speculate that the Action Observation Network (AON) might not even have any functional role in this regard (Hickok, 2009). They further claim that such confusion arises from the lack of specificity of the ‘action understanding’ concept used in the literature. Indeed, actions can be described at multiple levels and, consequently, there are multiple levels at which an observed action can be ‘understood’. Following this, they postulate that hierarchically distinct cerebral regions (having ‘mirror’ properties or not) subserve the different levels of action understanding: i) the *kinematic* level, which refers to the trajectory and velocity profile of the observed goal-directed action, including both the reaching and grasping phase of the action; ii) the *motor* level, which refers to the processing and pattern of muscle activity required to produced the desired kinematics; iii) the *goal* level, which refers to the immediate purpose of the observed action, such as grasping an object; and iv) the *intention* level, wich refers to the general (i.e., super-ordinate) reason, or cause, that motivates the execution of the observed action (see figure 5).

The ‘predictive coding’ framework postulates that each level of the hierarchy generates predictions that specify, or bias, the representations at the inferior level. Predicitions generated at the highest levels are then compared with current predictions generated at the inferior levels. Thus, depending on the type of goal being anticipated, the observer will predict a motor command that is congruent with this goal and, on the basis of her own motor repertoire, will generate expectations about the specific kinematics that best fits with the predicted motor command. The comparison between the predicted and the currently observed kinematics will generate, in the case of a mismatch, a prediction error. The magnitude of such a prediction error will then be reduced, via *forward* connections, by updating the representation elaborated at the more abstract, higher levels. This exchange of reciprocal signals proceeds until the cause that most likely explains the observed action is inferred, or, in other words, until the magnitude of the prediction error is sufficiently minimized to enable accurate inference (Kilner et al., 2007a; 2007b).

In this respect, the ‘predictive coding’ model provides a biologically plausible framework for explaining how the brain solves the inverse problem of action understanding by minimizing, at each cortical level (kinematic, motor, immediate goal, and intention), the prediction error that is automatically generated during action observation (see figure 6). When extended to a hierarchical model of brain functioning, such predictive coding also accounts for how top-down influences can be produced in a dynamic and context-specific way, through internal properties of the system itself (Friston et al., 2006). In this context, the estimations generated at the superior levels are the “priors” that inform the inferior levels. By potentiating sensory information via top-down adjustments of the prediction error signal, these high-order estimations make possible intention inferences, even in the case of noisy or incomplete perceptual information (Baker et al., 2006; Summerfield and Egner, 2009).

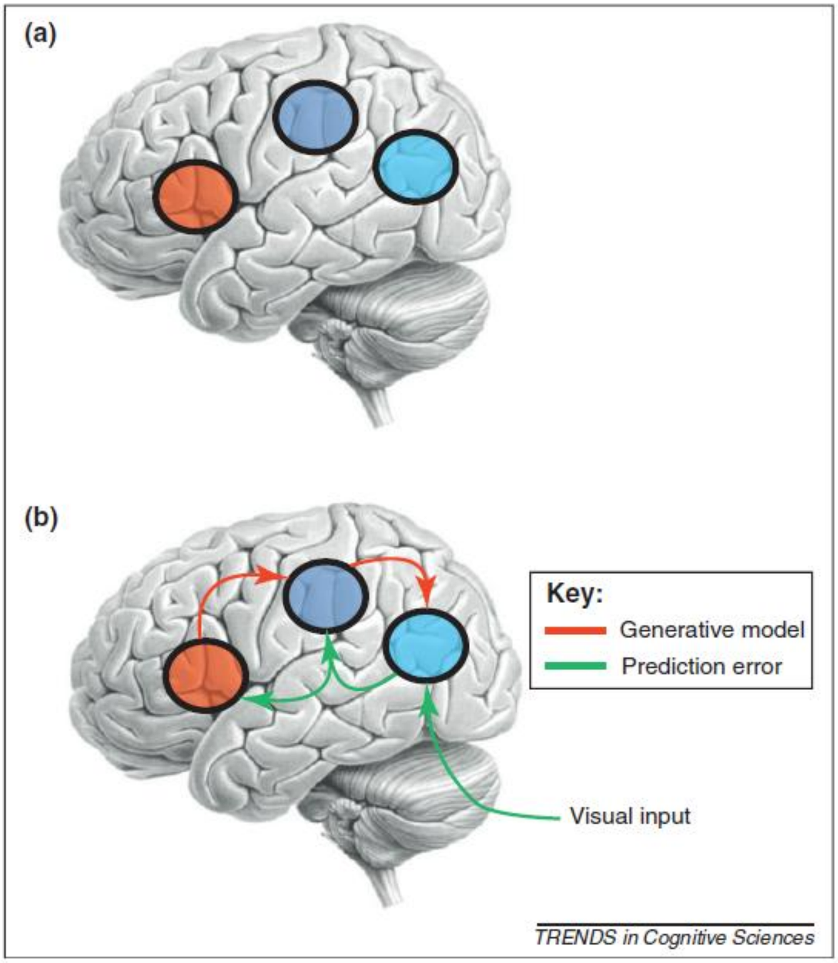


Figure 6. (a) Representation of the three reciprocally connected areas of the human 'mirror system', also called the Action Observation Network (taken from Kilner 2011). Some frontal (such as the ventral IFG) and parietal (such as the inferior parietal area) regions are known to contain mirror neurons. These frontal and parietal areas are reciprocally connected (Luppino, 1999) creating a premotor-parietal mirror system. Neurons within the STS have also been shown to respond selectively to biological movements, both in monkeys (Oram and Perrett, 2004) and in humans (Frith and Frith, 1999; Alisson et al., 2000; Grossman and Blake, 2002). The STS is reciprocally connected to the inferior parietal area (Seltzer and Pandya, 1994; Harries and Perrett, 1991) and therefore provides visual input to the mirror system. **(b) Representation of the predictive coding model of the AON (taken from Kilner, 2011).** Predictive coding is based on minimising prediction error though recurrent or reciprocal interactions among levels of a cortical hierarchy. In the predictive coding model, at each level of a cortical hierarchy a generative model is implemented which predicts the representations in the level below. This generative model uses backward connections to convey the prediction to the lower level. This prediction is then compared to the representation in this subordinate level to produce a prediction error. This prediction error is then sent back to the higher level, via forward connections, to tune the neuronal representation of sensory causes, which in turn changes the prediction.

2. Rethinking the function of the AON: two pathways to action understanding

Recently, Kilner (2011) proposed that, at the cerebral level, the concrete and abstract features of observed actions are encoded through a dorsal and a ventral pathway. A dorsal pathway – the AON –encodes the more concrete levels of actions (the kinematic and the motor levels), while the encoding of the more abstract levels (the immediate goal and intention levels) depends on the interaction between the AON and a ventral pathway linking the middle temporal gyrus (MTG) with the anterior inferior frontal gyrus (IFG).

The AON involves three main regions that are reciprocally interconnected. Two of these regions – the inferior parietal lobule (IPL) and the inferior frontal cortex (IFC) – are endowed with mirror properties. The IFC and the IPL would be involved in the coding of action goals. The more posterior parts of the IFC – the dorsal (PMd) and ventral premotor cortices (PMv) – are involved in the selection of the goal representation that best matches the observed action, Other parietal regions located near the IPL also participate in decoding features of an observed action. In particular, the anterior intra-parietal sulcus (AIPs) is known to be involved in the processing of manual prehension movements of object-directed actions, and is assumed to code for object/goal interactions (Tunik et al, 2007). These two clusters of brain regions receive inputs from the superior temporal sulcus (STS), which contains polysensory neurons that respond to motion from different perceptual modalities (Barraclough et al., 2005). The posterior part of the STS is potentially involved in the identification of intentional biological movements (Van Owervalle and Baetens, 2009).

According to the predictive coding model, action understanding would, and could, not only rely on the sole Action Observation Network. Rather, this network would be informed by prior expectations about goals and intentions that are formed along the ventral pathway. When the most *likely* goal has been specified, the sensory consequences of the action would be inferred by the dorsal pathway through brain areas of the AON. Note that prediction of the most likely intention could possibly be estimated outside the AON, from an analysis of contextual information surrounding the action scene (possibly through regions of the ‘mentalizing’ system such as the anterior frontomedian cortex, (Brass et al., 2007).

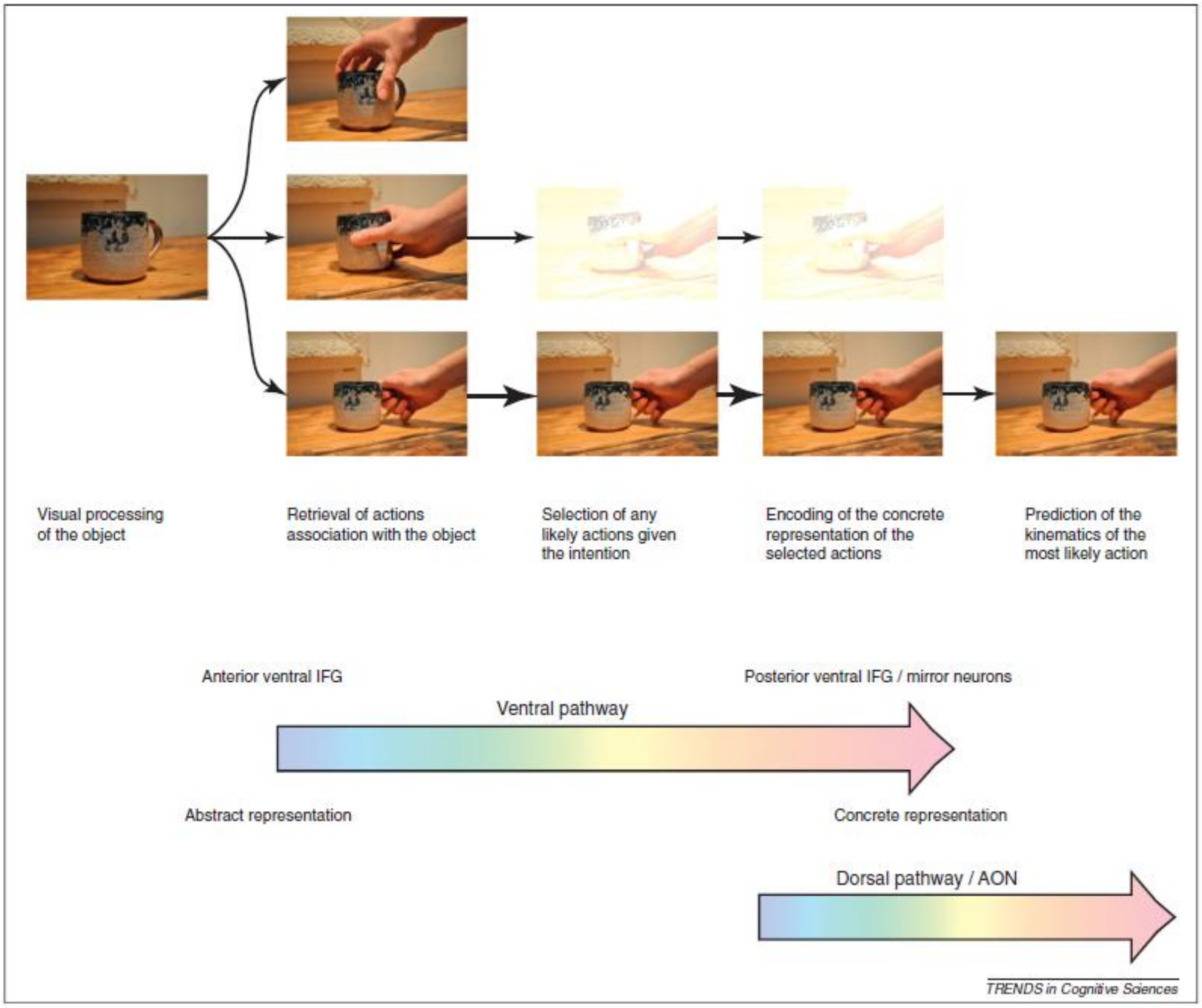


Figure 5. Example of action understanding in the two-pathway framework (taken from Kilner, 2011). The demonstrator's intention is to drink a cup of tea. The first step of action understanding is the visual processing and the identification of the object as a cup. The second step is the retrieval, within the observer's motor repertoire, of actions that he learned to be associated with that object. The third step is the selection of the most probable actions that can be achieved, given the demonstrator's final intention. Of note is that many actions can be selected. However, the likelihood of an action is signalled through the strength of that action's representation (indicated by the transparency of the picture). The top action is less probable and thus is not selected. The fourth step is the encoding of the motor parameters to generate a prediction of the sensory consequences of the observed action. Again multiple actions can be encoded as before. The fifth step is the prediction of the sensory consequences of the most probable action. Here only the most probable action is encoded. In this schematic, steps 2–4 would be encoded in the ventral pathway of the connected areas MTG, BA47, BA45 and BA44/BA6 with the representation of the action changing from the abstract to the concrete through these steps (left rainbow arrow). Steps 4–5 would represent the generation of the predicted sensory consequences of the action encoded in the dorsal AON pathway (right rainbow arrow).

F. Experimental outlines

The experiments presented in this doctoral thesis are guided by a series of questions that can be articulated along two dimensions: an individual dimension, through which processes of action recognition and action prediction are investigated; and a populational dimension, through which mechanisms of social transmission are modeled to account for the emergence of behavioral patterns among groups of individuals.

1. The first study (**Experimental Contribution Chapter 1**) concerns the representational aspect of perceived actions. More specifically, it aimed to investigate whether, and how, the inferior frontal cortex (IFC), the anterior intraparietal region (AIP) (two regions composing the Action Observation Network), and the primary somatosensory cortex (S1), are involved in the coding of two specific action features that are hierarchically organized. The first feature is the type of grip used by an observed agent to perform a specific action. The second feature is the subsequent arm movement that leads to the achievement of the action outcome. To do so, a simple action recognition task was designed. In a first step, participants were required to observe repeated movies showing a demonstrator acting on a multipurpose tool in order to achieve one out of two different goals by using either a power or a precision grip. In a second step, static pictures presenting the action outcomes were shown, and participants were asked to match either the current grip or the current arm action with the one performed in the preceding action movies. To assess the role of the three mentioned target brain areas, a single-pulse

transcranial magnetic stimulation technique was applied during the presentation of each static picture.

2. The second study (**Experimental Contribution Chapter 2**) was built upon the assumption that action prediction amounts to solving an inverse problem. This study thus aimed to investigate whether such a problem could be solve through a simple heuristic – such as the detection of object affordances – or required a more complex form of inference – such as probabilistic inference. More specifically, we investigated here the contribution of two types of prior information to the prediction of actions presented under various conditions of visual uncertainty. Using typical techniques from experimental psychology, we designed an action prediction task in which participants were required to infer the intentions of a filmed demonstrator acting on a multipurpose tool by using either a biomechanically *optimal* or *suboptimal* strategy. The first type of prior information that was manipulated directly emerged from the detection of object affordances, and provided expectations about the biomechanical optimality of the observed actions. The second type of prior information that was manipulated was the prior knowledge that observers derived about the demonstrator's behaviors and intentions from probabilistic sampling of past observations. Inspired from the framework of Bayesian inference, this experiment studied whether, and how, these two types of priors – biomechanical and probabilistic priors – interact to actively bias predictions made by participants.

3. The objective of the third study (**Experimental Contribution Chapter 3**) was to investigate whether, and how, the interaction between biomechanical and probabilistic priors

modulates the activity of the motor system in a task where participants were required to predict a demonstrator’s intentions. To do so, we adapted experiment #2 to an on-line TMS design. Corticospinal excitability was measured by means of single-pulse TMS applied over the primary motor cortex (M1) during action prediction, before and after probabilistic exposure to optimal and suboptimal actions.

4. The fourth and last study (**Experimental Contribution Chapter 4**) tested the extent to which some constraints, such as the richness of the biomechanical repertoire of an individual, or the type of search space characterizing a problem (two issues that relate to the notion of inverse problem), may influence the emergence of faithful social learning strategies (e.g., emulation or imitation) at a population level, and whether the emergence of such strategies may explain the stability of observed behavioral traditions. To explore these relationships, an evolutionary individual-based model was built.

EXPERIMENTAL CONTRIBUTION

CHAPTER 1

The first experiment presented in this thesis focused on a perceptual aspect of action understanding. In particular, it aimed at studying whether, and how, the primary somatosensory cortex (S1) – together with the inferior frontal cortex (IFC) and the anterior intraparietal region (AIP) that are classically reported as forming the Action Observation Network (AON) – is involved during the perception of distinct features of complex actions, i.e., the type of grip used and the type of effector configuration (arm) associated with action outome. Indeed, recent studies suggested that S1 may be recruited during the perception of observed actions within an estimation of their sensorimotor and somatic consequences (Keyser and Gazzola, 2007; Keyser et al., 2010; Valchev et al., 2012).

To investigate the respective contribution of S1, AIP and IFC in the perception of these action features, we used a transcranial magnetic stimulation adaptation paradigm (TMSA) (Silvanto et al., 2008). The underlying assumption of TMSA paradigms is that the impact of magnetic stimulation over a neural population does not only depend on the stimulation properties itself, but also on the initial state of the neural population prior to delivery of the stimulus. TMS may thus have a differential effect on neurons according to their initial activation state. Specifically, TMS behaviorally facilitates the detection of perceptual features enconded by adapted neural populations, compared with non-adapted, within the stimulated brain area (Silvanto et al., 2008). TMSA paradigms consist in manipulating the initial state of neural populations prior to the stimulation by perceptual adaptation. The adaptation phase simply consists in exposing the participant to the repetition of a stimulus (e.g., a visual stimulus) for a duration usually comprised between 40and 60 sec. The adaptation induces habituation in a subset of cells that code for a particular feature of that stimulus, making them a selective target for TMS. Consequently, stimulation time locked to the cognitive task (e.g., the detection of the adapted stimulus among non-adapted stimuli) should selectively improve the performance in processing the stimulus to which the targetd neurons were previously adapted. TMS adaptation paradigms have been recently successfully tested in various domains of cognition, such as language (Cattaneo et al., 2009a; 2009b) and number processing (Cohen Kadosh et al., 2010), multisensory interaction (Romei, 2007) or motor acts observation (Cattaneo et al., 2010; 2011).

In line with these works, we used a TMSA paradigm in two tasks of action recognition. During these tasks, participants were habituated to adaptation movies showing a human actor performing two types of complex object-directed actions (lifting the object to open a box versus turning the object to switch-on a light) using either a power (whole hand grip) or a precision grip (pinch grip). After each adaptation phase, a series of static pictures presenting the final end-states of each possible possible action was presented. In the first task, participants were required to detect the similarity between the grip currently used in the static pictures and the grip previously used in the adaptation movie. In the second task, they were required to detect the similarity between arm movements that led to the final action end-states of the static picture and those observed in the preceding adaptation movie. At the onset of each static picture, single-pulse TMS was delivered over the IFC, S1 and AIP. We aimed at investigating the role of these structures in the coding of two key features of actions, namely the type of grasp and the type of effector configuration that leads to the outcome of the observed action.

Perturbing the Action Observation Network during the perception and categorization of others' actions: state-dependency and virtual lesion TMS effects.

(Preliminary draft)

Pierre O. Jacquet[1,2], and Alessio Avenanti[1,3]

[1] Department of Psychology, University of Bologna, Alma Mater Studiorum, Bologna, Italy

[2] INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Center, ImpAct Team, Lyon, France

[3] Center for the Study and Research in Cognitive Neuroscience, Scientific and Didactic Pole of Cesena, Cesena, Italia

Abstract

The perception of other people's actions is negociated by a large cerebral network distributed along a parieto-frontal axis – the Action Observation Network (AON). The AON, notably composed of the anterior intraparietal area (AIP) and the inferior frontal cortex, is assumed to cods action components into a hierarchical manner, ranging from movement kinematics to action goals and intentions. Recent studies suggest that other sensory areas, such as the primary somatosensory cortex (S1), could actively participate in action perception. Yet, no causative evidence has been brought in favor of such involvement. Furthermore, it is not yet clear whether, and how, the AIP, the IFC and the S1 cortices are differentially recruited during the perception of observed complex goal-directed actions. The objective of the present study is to investigate these issues by using a transcranial magnetic stimulation adaptation technique (TMSA). TMSA paradigms allow to scrutunize the causal involvement of specific brain areas in the achievement of behavioral tasks, and so by manipulating the state of these regions through the adaptation of neuronal populations to a constant stimulus. Here, we used single-pulse TMS over the left AIP, IFC and S1 cortices after the exposure to complex goal-directed actions. Participants repeatedly observed adaptation movies in which a demonstrator acted on a two-purpose tool by achieving two distinct action goals using either a power or a precision grip. After the exposure to adaptation movies, test pictures of the four possible action end-states (2 grips × 2 intentions) were presented. In a first task, participants were required to judge whether the test pictures presented a similar or a different grip compared to the one previously observed during the adaptation movies while in a second task, they were asked to judge whether the type of arm action used by the demonstrator to achieve his goal was similar or different. TMS was applied at the onset of each test pictures over the AIP, IFC and S1 cortices. If these regions play a role in the processing of the grip or the arm action features, the recognition of the adapted stimulus should be improved. First, results showed that stimulating the IFC impaired the recognition of the grip, independently of its type (adapted or non-adapted), and impaired the recognition of the non-adapted arm action. Second, stimulating S1 improved the recognition of the adapted arm action. These findings provide the first evidence that the IFC and the sensorimotor regions of the AON such as S1 are causatively involved during the perception of complex-goal directed actions.

Keywords: action perception, mirror system, state-dependency TMS, Action Observation Network, primary somatosensory cortex.

Introduction

Interpreting the behaviors of others crucially requires the ability to decode the observed kinematic information into a causal relationship between the motor sequence and its outcome. It has been suggested that this ability depends on action ‘simulation’ mechanisms mapping observed motor acts onto corresponding motor programs and associated somatosensory consequences stored into the observer’s action repertoire (Rizzolatti and Craighero, 2004; Wilson and Knoblich, 2005; Urgesi et al., 2010; Keysers et al., 2010; Avenanti and Urgesi, 2010). Imaging and neurophysiological evidence have suggested that such action simulation mechanisms rely on the activity of a widespread bilateral network of cortical brain regions, usually referred to as the action observation network (AON) (Grafton, 2009). Classically, inferior frontal cortex (IFC, including the ventral premotor cortex and the posterior part of the inferior frontal gyrus) and the anterior intraparietal cortex (AIP) have been considered important nodes of the AON coupling action observation with execution. Seminal studies on monkey indicate that a proportion of neurons in these frontoparietal regions increase their firing rate during both action perception and execution (so called ‘‘mirror neurons’’) (di Pellegrino et al. 1992; Gallese et al. 1996; Fogassi et al. 2005) and may be involved in perceiving and understanding others’ actions.

In addition, mounting imaging and neurophysiological evidence in humans suggest that the somatosensory cortices may be also involved in perceiving others’ behavioral states (Adolphs et al., 2000; Keysers et al., 2004; Bufalari et al., 2007; Valeriani et al., 2008; Avenanti et al., 2009). In particular, the primary somatosensory cortex (S1) is consistently active during action perception and execution (Rossi et al., 2002; Avikainen et al., 2002;

Avenanti et al., 2007; Gazzola and Keysers, 2009) and may be thus be considered an additional node of the AON (Keysers et al., 2010).

Although the involvement of IFC, AIP and S1 in perceiving others' actions has been suggested in several imaging and neurophysiological studies (Caspers et al., 2010), it should be noted that these methods provide only correlational evidence and cannot establish a direct causal link between brain and function. Transcranial magnetic stimulation (TMS) provides an extraordinary non-invasive method to interact with neural tissue and thus provide that a brain region is critical for behavior. Typically, TMS is used with the aim of disrupting neural activity associated with cognitive processes by inducing random neuronal activity that is uncorrelated with the ongoing activity (i.e. "virtual lesions"). Previous studies using these methods have suggested that disruption of activity in IFC (Pobric and Hamilton, 2006) or S1 (Valchev et al., 2012) reduces the ability to infer the weight of objects being lifted by a human hand. Moreover, inhibition of IFC has been shown to generate several disturbances affecting various aspects of action perception and understanding. In particular, inhibition of IFC can result in the disruption of action simulation activity within the motor system (Avenanti et al., 2007; Avenanti et al., 2012), the impairment of the ability to discriminate between two different pictures of actions (Urgesi et al., 2007; Candidi et al., 2008), or the impairment to discriminate between deceptive or truthful actions (Tidoni et al., 2012, unpublished observations). These virtual lesions studies indicate that manipulation of neural activity in the AON (mostly in the IFC) impairs action perception. Critically however, to date no studies have compared the causative influence of IFC, AIP and S1 in an action recognition task.

Moreover, virtual lesions approaches suffer from two main limitations. First, the effect of brain stimulation is not limited to the target brain region, but can spread ortho- and antidromically along neural connections (Avenanti et al., 2007; Avenanti et al., 2012). Hence, TMS may be best conceptualized as modulating activity across large scale networks reached from the directly targeted brain region (Valero-Cabré et al., 2005, 2007). Second, even within the targeted brain region, virtual lesion TMS approach cannot elucidate how distinct neural populations within the stimulated area interact to give rise to perception and behavior (Silvanto and Pascual-Leone, 2008; Avenanti and Urgesi, 2010).

Recently, TMS-adaptation (TMSA) paradigms have been developed to tackle these limitations and to provide information on the cortical topography of brain functions and the causal relation of neural activity in the targeted areas to behavior. The TMSA paradigm is based on the well established notion of state-dependency, i.e. that TMS effects depends on the context and the initial state of the stimulated neurons. Specifically, TMS is thought to differentially modulate neurons that are activated by a given perceptual or cognitive process relative to neurons that are not activated by such process (Silvanto et al. 2008). Evidence suggests that TMS preferentially facilitates the less active neurons and/or suppresses the more active neurons within a stimulated brain region and it has been shown that this differential modulation have behavioral consequences (Cattaneo and Silvanto, 2008).

In the TMSA paradigm, the state of the neurons prior to the TMS pulse is manipulated in a controlled way by means of perceptual adaptation. The adapting stimulus induces habituation in a subset of neurons that code particular stimulus features, making them a selective target for TMS. Stimulation time-locked to the cognitive task and delivered over the cortical area containing the adapted neurons should selectively improve the performance in processing the adapted stimulus features relative to the non-adapted ones and/or reciprocally, should impair the performance in processing the non-adapted stimulus feature relative to the adapted ones. Hence, TMSA method allows targeting functionally distinct but spatially overlapping neural populations (Silvanto et al., 2009). This paradigm has been successfully used to explore color and motion perception in the visual cortex (Silvanto and Muggleton, 2008) as well as to investigate language and number processing in the parietal cortex (Cattaneo et al., 2009; Cohen-Kadosh et al., 2010). However, it should be noted that brain stimulation during TMSA paradigm may also lead to disruption of neural function i.e. to “state-independent” virtual lesions effects (Burton et al., 2009).

Recently, the TMSA technique has been used by Cattaneo and colleagues to explore action perception mechanisms (Cattaneo et al., 2010a). In this study, participants observed adapting movies showing an actor performing grasping or pulling motor acts with either the hand or the foot. After each series of adapting movies, participants were subsequently presented a series of test pictures showing the same or different motor acts, matched for effectors and/or actions. For each test pictures, they were asked to provide a similarity judgment, i.e. whether the current picture presented an action that was identical or different to the one presented in the preceding adapting movies, irrespective of the effector used to perform the action. It was found that response times (RTs) to adapted actions were shorter when TMS was applied over two visuo-motor nodes of the AON, namely the IFC and AIP. This improvement was independent of the type of effector that was involved in the adapting movies, while stimulation of a visual node of the AON (the superior temporal sulcus, STS) led to effector specific improvements. These findings suggested a hierarchy in the representation of others’ actions, with visual and visuo-motor nodes of the AON being critically involved in processing others’ actions in an effector specific and more abstract manner, respectively. However, in that study, RTs and accuracy measures were analyzed separately and thus it is not clear whether speed accuracy trade-off effects were at play in the study. Moreover, no sham stimulation condition was present in the design and thus unspecific effects of TMS are not determined.

The experiments performed in the present study are inspired from the Cattaneo et al.'s paradigm (2010a), but proposed a finer-grained analysis of action perception mechanisms. Indeed, instead of manipulating the type of effector (i.e., hand versus foot), we kept the same effectors (i.e., the hand) but manipulated the type of action the actor could perform (i.e., lifting an object to open a box versus turning an object to switch-on a light; see fig.1) and the type of grip the actor could use to perform the action (i.e., power versus precision grip). Critically, subjects had to perform two tasks in which they had to provide similarity judgments both on the arm action (irrespective of the grip being used to perform the action; Arm action recognition task) and on the grip used in the action (irrespective of the type of action; Grip recognition task) while active TMS was applied to the left IFC, left AIP and left S1, and sham TMS was applied to the vertex. To rule out any speed-accuracy trade off, the effect of TMS on the two action perception tasks was assessed by merging both RTs and accuracy into a single measure of performance (inverse efficiency index). By this way we were able to test the causal involvement of key nodes of the AON in the differential visual coding of specific actions (lifting vs opening) and action components (power versus precision grip).

Materials and Methods

Participants

Twenty-seven healthy adults (16 females, 11 males; mean age = 25.8, SD = 5.18) volunteered to participate in the study. All participants were right-handed, reported normal or corrected-to-normal vision acuity, and were naïve to the purpose of the experiment. None of them reported evidence for neurological or psychiatric diseases. The experimental setting was approved by the ethics committee of the Bologna University's Department of Psychology and was carried out in accordance with the ethical standards of the 1964 Declaration of Helsinki. All participants gave a written and informed consent and were remunerated 24 Euros for participating in the study.

Stimuli

Participants were seated in a comfortable chair in front of a 19-inch computer screen on which series of video clips (adaptation movies) and static pictures (test pictures) were displayed. Adaptation movies consisted in 2000ms video clips (30 frame / second, subtending 35 degrees of visual angle) showing a male actor operating on an unfamiliar tool. The tool consisted of a handle designed to make possible two different goal-oriented actions using two different types of hand grip. More specifically, the actor could either i) lift the handle to open the box; or ii) turn the handle to switch-on the light. These two actions could be performed using either *e)* a power (whole-hand) or *ee)* a precision (pinch) grip (see figure 1). Thus, four types of videos were created following a 2 (type of action: lifting, turning) × 2 (type of grip: power, precision) design. A total of 24 clips for each category were created. All movies were equalized for temporal homogeneity such that the initial (from still hand to tool grasping) and last phase of the action (tool manipulation) involved the very same number of video frames and lasted 1000 ms each. In order to minimize the influence of memorized kinematics on the participants' performances, a movie was presented only one time per adaptation phase.

Test pictures consisted in single frames extracted from the four possible adaptation movies. They presented the actor in the four possible postures, corresponding to the final end-state of the four possible motor actions (i.e., lifting with power grip, lifting with precision grip, turning with power grip, turning with precision grip). Each test picture lasted 1500ms.

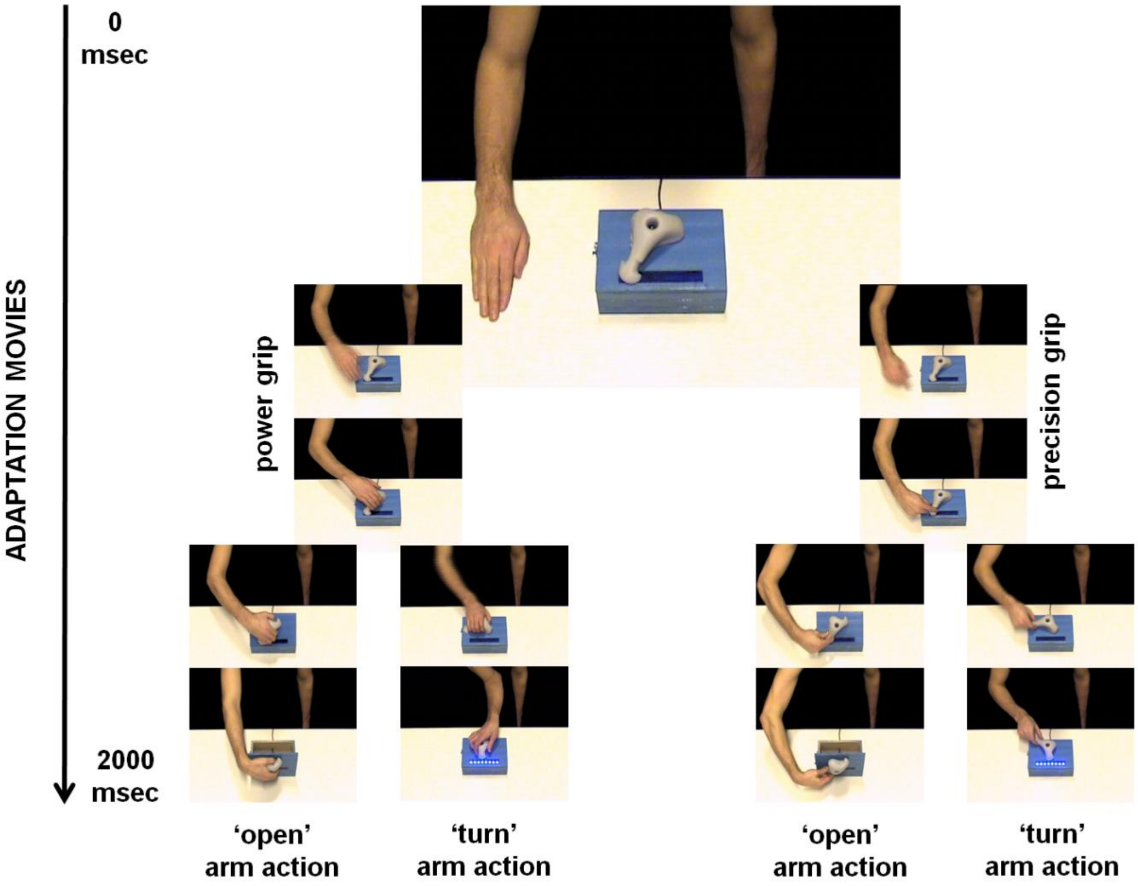


Figure 1. Example of the four adaptation movies that participants encountered during the experiments. All adaptation movies had a duration of 2000 msec and all began with the demonstrator's static hand. The actor could then use either a 'power' or a 'precision' grip to achieve either the action of lift the object to open the box ('open' arm action) or turning the object to switch on the light ('turn' arm action).

Tasks

Subjects performed two tasks (Action recognition, Grip recognition) in two separate sessions on the same day. In both tasks, subjects were presented with a series of adaptation movies showing one of the four motor actions (lifting with power grip; lifting with precision grip; turning with power grip; turning with precision grip) followed by a series of 12 test pictures (3 exemplars for each action). In the Arm action recognition task, participants were asked to identify, for each test picture, whether the action (lift versus turn the handle) implied in the picture was similar or different as compared to that shown in the preceding adaptation movies (independently of the grip used in the action). In the grip recognition task, they had to reported whether the grip depicted in the picture (power versus precision grip) was similar or different relative to that shown the preceding adaptation movies (independently of the type of action being performed). Responses were made with the index and middle finger of the left hand (ipsilateral to the stimulated hemispheres) on a keyboard and responses time (RTs) and accuracy were recorded and analyzed off-line.

General Procedure

Participants performed the two tasks in two sessions separated by a pause of 15min duration. The order of the tasks was counterbalanced across subjects. The Eprime 2.0 software (Psychology Software Tools, Inc, USA) was used to collect both reaction times (RTs) and percentages of correct responses.

Action and Grip recognition tasks were performed in 4 different TMS blocks (Sham, IFC, S1, AIP) whose order was randomized across subjects. Each block included 4 trials. On each trial, participants were first presented with 30 adaptation movies displaying one of the four possible motor actions. Movies were separated by a white fixation cross of 250ms duration. The series of adaptation movies (lasting 67.5 s in total) was followed by a written signal of 2000ms duration. This signal simply consisted in a pattern of 8 exclamation points the aim of which was to inform participants about the up-coming presentation of test pictures. Then a sequence of 12 test pictures (4 action types × 3 different exemplars) was shown. Test pictures lasted 1500 ms each and were separated by a fixation cross of 1000 ms of duration. At the onset of each test picture, a single-pulse TMS was delivered (Silvanto et al., 2007; Cattaneo et al., 2010a). Participants were first required to carefully watch the series of adaptation movies and then they had to compare the test pictures to the movies (see figure 2). A fixation cross was shown in the inter-trial interval (15 s duration). The order of the trials and of the test pictures was randomized. A total of 48 responses were then collected in each block (4 trials × 12 test pictures). In both tasks, half of the test pictures showed adapted Action/Grip configurations and half showed non-adapted Action/Grip configurations.

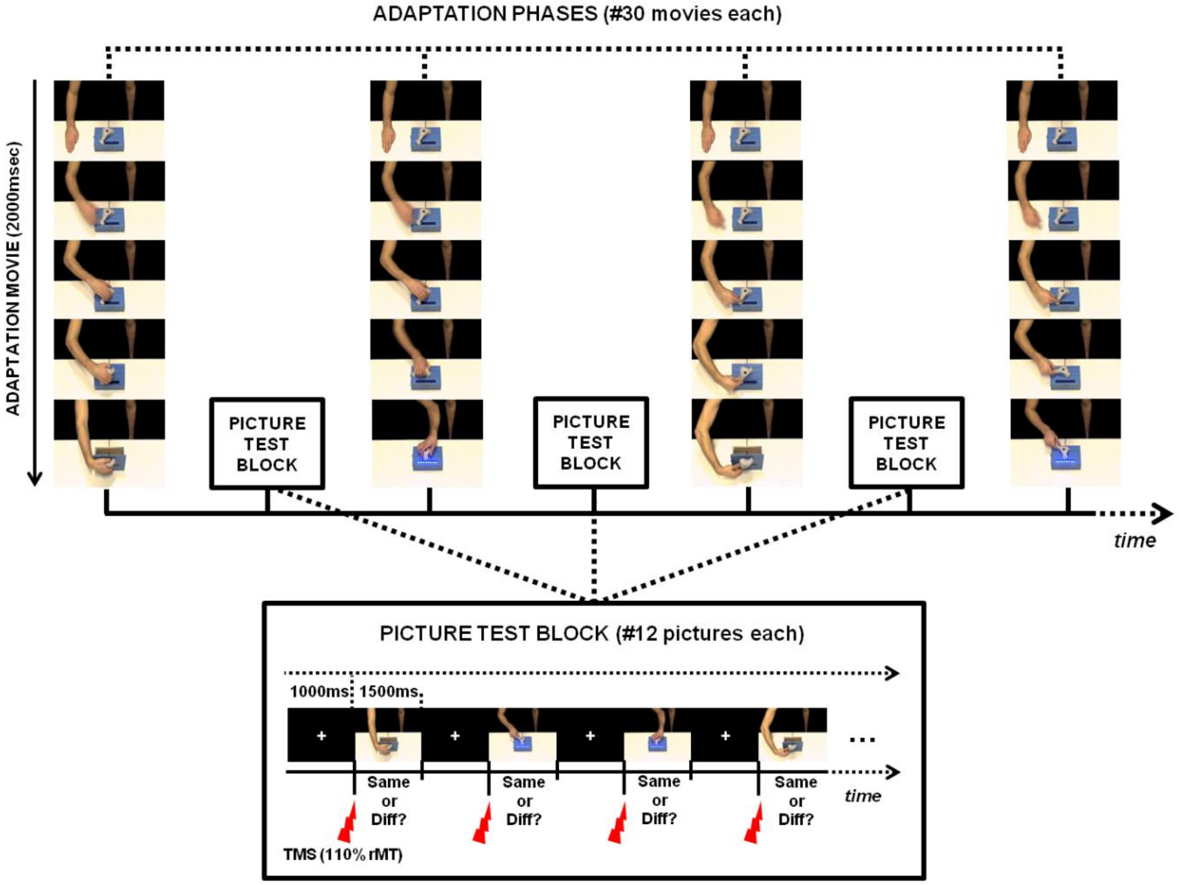


Figure 2. Example of a TMS block (each subject parformed 4 different TMS block: sham, IFC, S1, AIP). A TMS block included 4 trials. On each trial, participants were first presented with 30 adaptation movies showing one of the four possible motor acts (‘adaptation phase’). The series of adaptation movies was followed by a 12 test pictures (‘test pictures block’ consisted of 4 action types × 3 different exemplars). Test pictures lasted 1500 ms each and were separated by a fixation cross of 1000 ms of duration. At the onset of each test picture, a single-pulse TMS was delivered. Participants were first required to carefully watch the series of adaptation movies and then they had to compare the test pictures to the movies. A total of 48 responses were then collected in each TMS block (4 trials × 12 test pictures).

Transcranial Magnetic Stimulation

TMS pulses were delivered with a figure-of-eight coil (70mm) and a Magstim Rapid^2 stimulator (Magstim, Whitland, Dyfed, U.K.). The individual resting motor threshold (rMT) of each participant was identified as the minimal stimulation intensity producing motor evoked potentials (MEPs) of at least 50 μV amplitude in the right first dorsal interosseous muscle (FDI), with 50% probability on at least 5 on 10 consecutive stimulations (Rossini et al., 1994). MEPs were recorded by means of a Biopac Student Lab MP36 electromyograph (Biopac Systems, Inc, U.S.A.). EMG signals were band-pass filtered (30 Hz–1000 Hz) and digitized (sampling rate at 5 kHz). Pairs of silver/silver chloride surface electrodes were placed over the muscle belly and over the first articulation of the right index finger, whereas a ground electrode was placed on the internal bone of the right elbow. The intensity of stimulation used during the experiments was then set at 110% of the individual rMT. The experiment was programmed using Eprime 2.0 software (Psychology Software Tools, Inc.) to control stimuli presentation and trigger the magnetic stimulator.

Coil position was identified on each participant's scalp using the SofTaxic Navigator system (EMS, Italy) or functional methods as in our previous TMS research (Avenanti et al. 2007; Avenanti et al., 2012; Bertini et al. 2010; Serino et al. 2011). Skull landmarks (nasion, inion, and two preauricular points) and about 60 points providing a uniform representation of the scalp were first digitized by means of a Polaris Vicra Optical Tracking System (NDI, Canada). Coordinates in Talairach space were automatically estimated by the SofTaxic Navigator from an MRI-constructed stereotaxic template.

Scalp positions corresponding to IFC and AIP were identified by means of the SofTaxic Navigator system. The IFC was targeted in the anterior-ventral aspect of the precentral gyrus (ventral premotor cortex) at the border with the pars opercularis of the inferior frontal gyrus (coordinates: x = -52, y = 10, z = 24), corresponding to Brodmann's area 6/44 (Avenanti et al. 2012; Mayka et al. 2006; Van Overwalle et al. 2009; Caspers et al. 2010; Urgesi et al. 2007). The AIP region was targeted in the anterior sector of the intraparietal sulcus (x = -40, y = -40, z = 45, corresponding to Brodmann’s area 40; Van Overwalle et al. 2009; Caspers et al. 2010).

TMS studies that successfully targeted the hand region in S1 positioned the coil 1–4 cm posterior to the motor hotspot (Balslev et al., 2004; Avenanti et al., 2007; Valchev et al., 2012). In keeping, S1 was indentified using a two steps procedure. We first localized the hand region in the motor cortex (corresponding to the optimal scalp position for evoking MEPs in the FDI muscle) and then moved the coil 2 cm backward. We assumed that from this position we could stimulate the hand region in S1 with minimum effects on M1. To test this assumption directly, we checked that TMS pulses at 110% rMT with the coil in the above position did not elicit any detectable MEP. The three identified sites (IFC, S1, AIP) were marked on the bathing cap with a pen. Then the neuronavigation system was used to estimate the projections of the scalp sites on the brain surface (see figure 3).

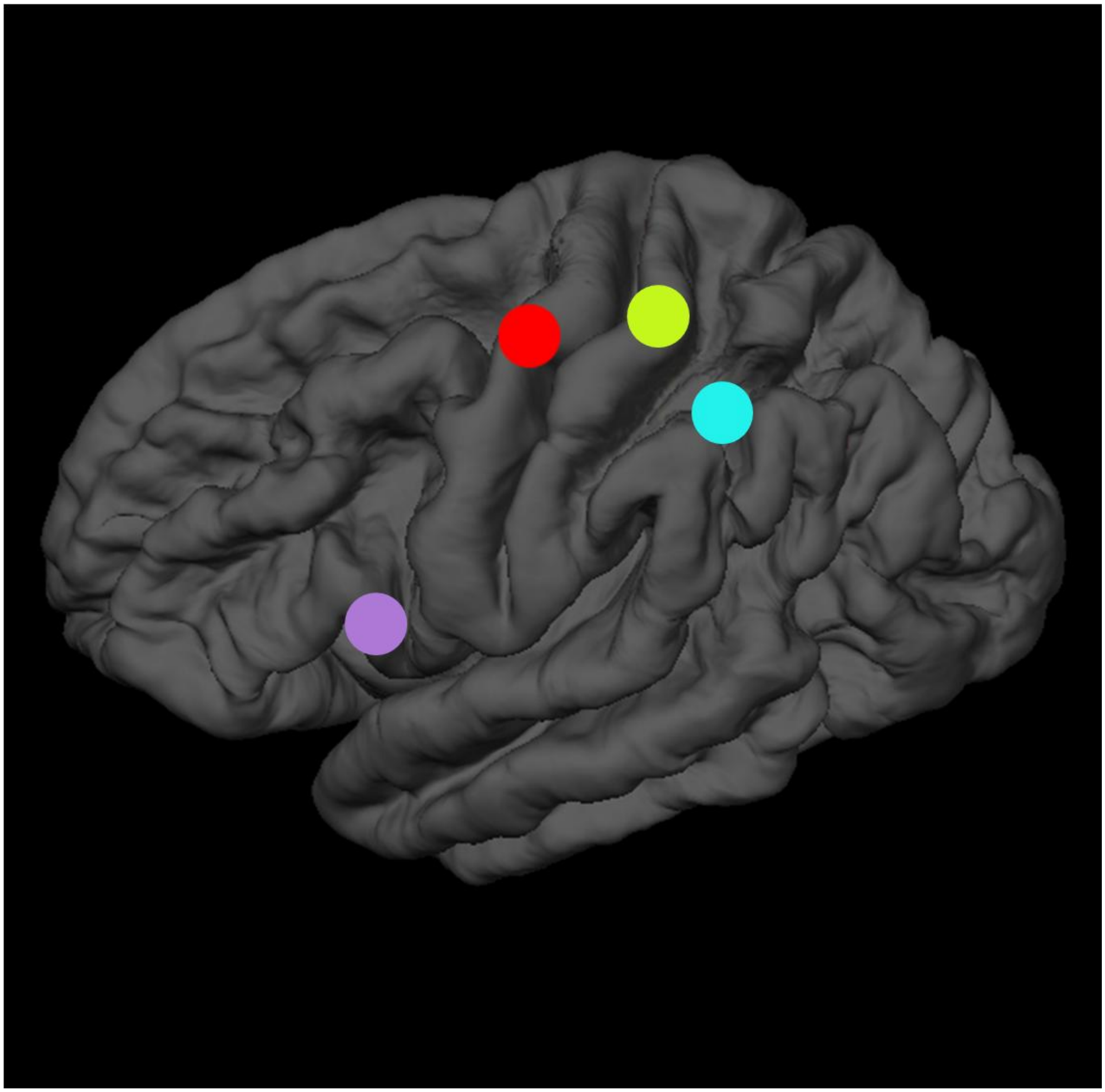


Figure 3. Schematic representation of the 3 stimulation sites. The red dot represents the cortical representation of the right FDI muscle located on the left prilary motor cortex (M1). It was used to individuate the resting motor threshold (rMT) on which the stimulator intensity was set (110% of the rMT). The purple dot represent the left inferior frontal cortex (IFC), targeted in the anterior-ventral aspect of the precentral gyrus (ventral premotor cortex) at the border with the pars opercularis of the inferior frontal gyrus (coordinates: x = -52, y = 10, z = 24). The green dot represents the left primary somatosensory cortex that was targeted by moving the coil 2 cm posterior to the FDI location. Finally, the blue dot represents the left AIP region, targeted in the anterior sector of the intraparietal sulcus (x = -40, y = -40, z = 45).

Data Analysis

The proportion of correct responses and RTs were analyzed off-line (see table 1 and 2). For each condition, RTs deviating more than two standard deviations from the individual mean were discarded. During the IFC stimulation, three participants reported being surprised by the stimulation in the first sub-block of test pictures (12 trials), resulting in an absence of response during this sub-block. This was actually a side effect of the stimulation which brought about facial muscle contractions and slight movements of the mandible. Accordingly, responses collected during this sub-block were removed from the analyses. No similar effects were found in the remaining blocks or in the other participants.

Statistical analyses were conducted on the inverse efficiency (IE) index (Akhtar and Enns 1989; Christie and Klein 1995; Kennett et al 2001; Townsend and Ashby, 1983), obtained by dividing the median RT by the proportion of correct responses, calculated for each experimental condition and for each subject separately. By combining response latencies and accuracy into a single measure, this index allows to discount possible criterion shifts or speed accuracy tradeoffs in the different TMS conditions. A lower value on IE indicates a better recognition performance, while a higher value on IE indicates a lower recognition performance. For the Arm action recognition task, IE scores were submitted to a 4 × 2 × 2 repeated-measures ANOVA with Stimulation (Sham, IFC, S1, AIP), Arm action (adapted, non-adapted) and Grip (adapted, non-adapted) as within-subject factors. A similar analysis was conducted on IE scores of the Grip recognition task. Before reporting the two ANOVAs, a preliminary Task (Action, Grip) × Stimulation × Grip × Arm movement ANOVA was carried out. Post-hoc analyses were carried out using the Newman-Keuls test.

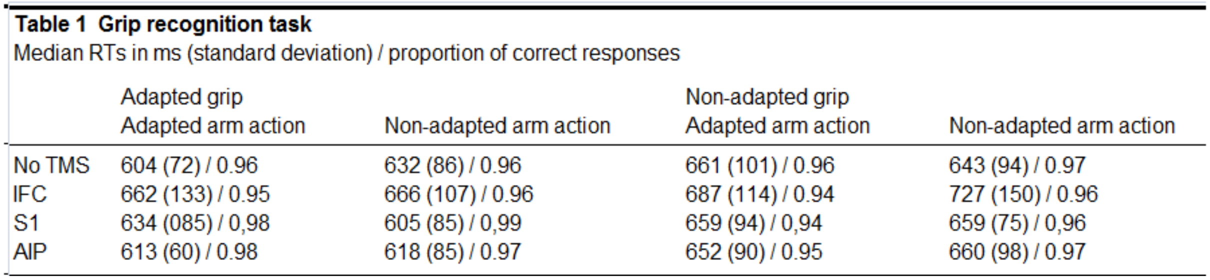


Table 1 Grip recognition task
Median RTs in ms (standard deviation) / proportion of correct responses

	Adapted grip		Non-adapted grip	
	Adapted arm action	Non-adapted arm action	Adapted arm action	Non-adapted arm action
No TMS	604 (72) / 0.96	632 (86) / 0.96	661 (101) / 0.96	643 (94) / 0.97
IFC	662 (133) / 0.95	666 (107) / 0.96	687 (114) / 0.94	727 (150) / 0.96
S1	634 (085) / 0,98	605 (85) / 0,99	659 (94) / 0,94	659 (75) / 0,96
AIP	613 (60) / 0.98	618 (85) / 0.97	652 (90) / 0.95	660 (98) / 0.97

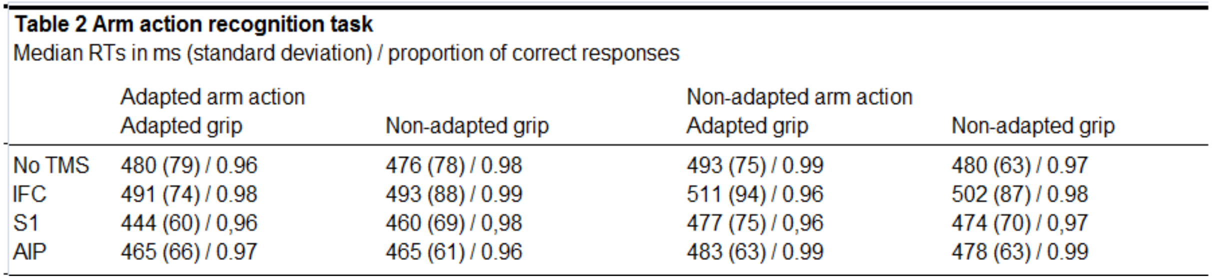


Table 2 Arm action recognition task
Median RTs in ms (standard deviation) / proportion of correct responses

	Adapted arm action		Non-adapted arm action	
	Adapted grip	Non-adapted grip	Adapted grip	Non-adapted grip
No TMS	480 (79) / 0.96	476 (78) / 0.98	493 (75) / 0.99	480 (63) / 0.97
IFC	491 (74) / 0.98	493 (88) / 0.99	511 (94) / 0.96	502 (87) / 0.98
S1	444 (60) / 0,96	460 (69) / 0,98	477 (75) / 0,96	474 (70) / 0,97
AIP	465 (66) / 0.97	465 (61) / 0.96	483 (63) / 0.99	478 (63) / 0.99

Results

The Task × Stimulation × Action × Grip ANOVA on the IE index revealed several effects including the main effect of Task ($F_{1,26}$ = 321.41, p < .00001) with greater IE values (lower performance) for the Grip (mean ± SD: 677 ms ± 76) relative to the Arm action recognition task (493 ms ± 64) and a marginally significant quadruple interaction ($F_{3,78}$ =2.67, p = .053). Thus, to further analyze the data, two separate Stimulation × Arm action × Grip ANOVAs were carried out, one for each Task.

Grip recognition task (see table 1)

The ANOVA on the IE index computed on the Grip recognition task revealed a main effect of Grip ($F_{1,26}$ = 20.30, p = .0001) with greater IE scores for non-adapted (701 ± 80) relative to adapted grips (654 ± 81) and a main effect of Stimulation ($F_{3,78}$ = 8.81, p = .00004), accounted for by the higher IE scores in the IFC block relative to the other blocks (all p < .0004) which in turn did not differ from one another (all p > .80) (see figure 5). No Grip × Stimulation interaction was found ($F_{3,78} = 1.57$, p = .20), suggesting that TMS did not induce any state-dependency effects in the Grip recognition task, but only a strong virtual lesion effect (lower performance) when applied over the IFC. No other significant main effects or interactions were found (all p > .18).

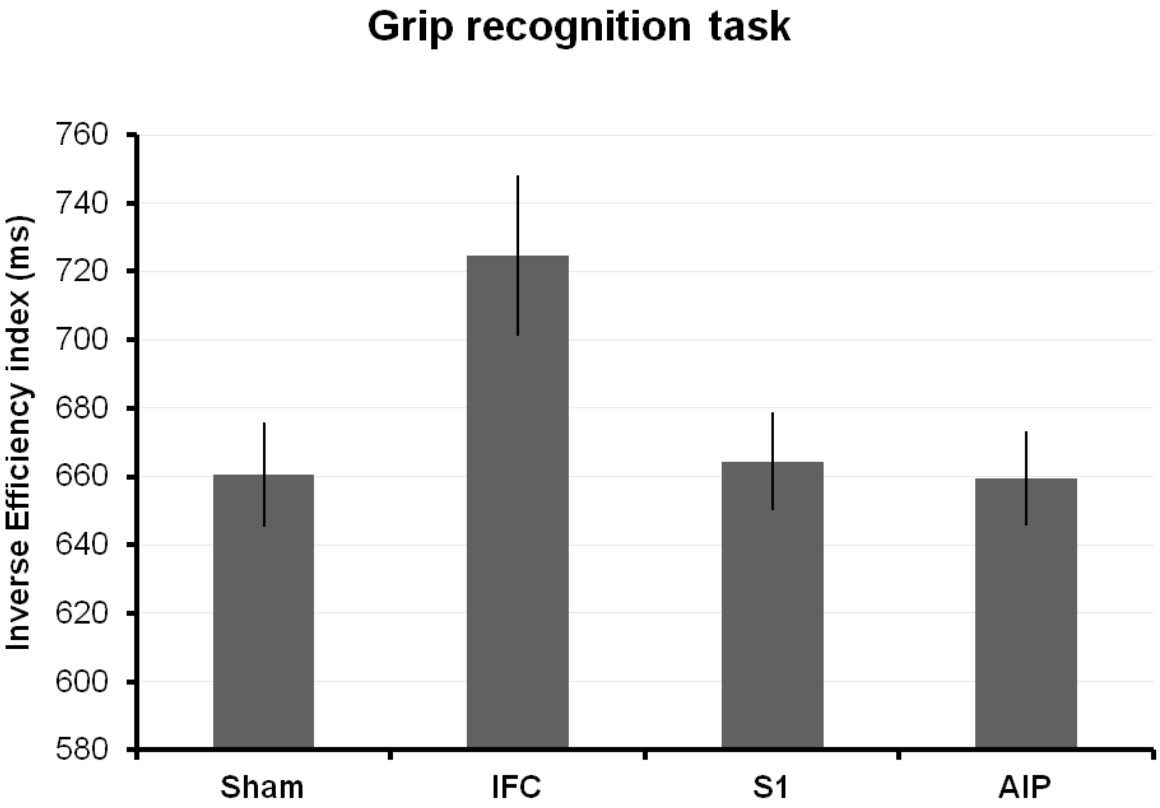


Figure 5. Arm action recognition task: interaction effect 'Stimulation' × 'Arm action'. The vertical axis represents the mean inverse efficiency (IE) index (obtained by dividing the median RT by the proportion of correct responses collected during the test pictures blocks of the Arm action recognition task). The horizontal axis represents the four stimulation condition. The black columns represent IE index calculated for the adapted arm actions. The grey columns represent IE index calculated for the non-adapted arm actions.
Arm action recognition task (see table 2)

The ANOVA on the IE index computed during the Arm action recognition task showed a main effect of Stimulation ($F_{3,78} = 4.27$, p = .0008), a main effect of Arm action ($F_{1,26} = 4.35$, p

= .047) and, most importantly, a significant Stimulation × Arm action interaction ($F_{3,78} = 3.12$, $p = .031$) (see figure 4). Post-hoc analysis (Newman-Keuls range tests) of the interaction suggested that no clear behavioral correlate of adaptation was present when Sham stimulation was administered as comparable IE scores were seen for adapted relative to non-adapted actions ($p = 0.48$). Similarly no difference between adapted and non-adapted actions were found in the AIP block ($p = 0.66$). In striking contrast, when IFC or S1 stimulation was administered a clear '*state change*' was induced in the subject's perceptual system as a function of prior exposure to the adapting movies. In the S1 block we found that adapted actions elicited lower IE (better performance) relative to non-adapted actions ($p = .004$) and to all the other conditions in the other blocks (all $p < .032$). In the IFC block, we found that non-adapted actions elicited higher IE (worst performance) relative to adapted actions ($p = .0005$) and to all the other conditions in the other blocks (all $p < .0006$). No other significant post-hoc differences were found ($p > .13$). Moreover, no other main effect or interactions was significant in the ANOVA (all $p > .18$).

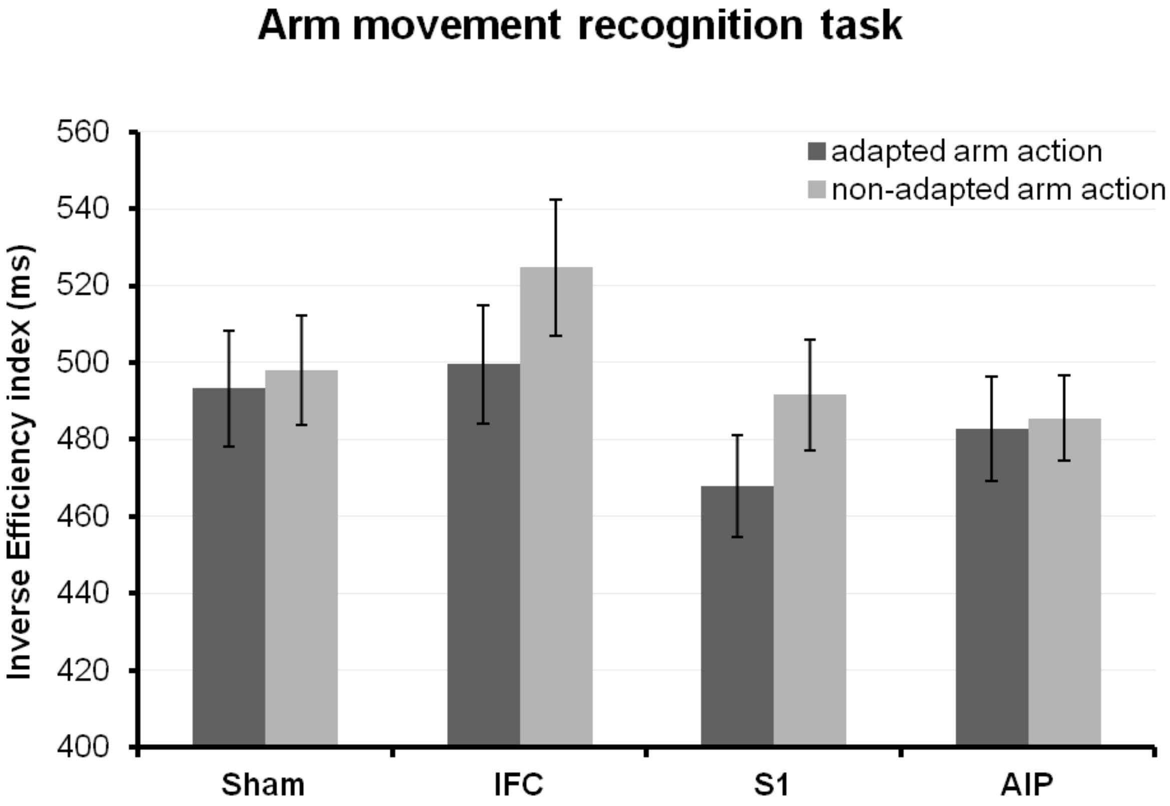


Figure 4. Grip recognition task: main effect of the 'Stimulation' factor. The vertical axis represents the mean inverse efficiency (IE) index (obtained by dividing the median RT by the proportion of correct responses collected during the test pictures blocks of the Grip recognition task). The horizontal axis represents the four stimulation condition.

Discussion

Mounting evidence suggest that the ability to recognize the actions of others is underpinned by a large cortical network, called the AON (Grafton 2009), which includes occipital and temporal regions involved in the visual processing of body and biological motions (Keysers and Perrett, 2004; Downing and Peelen, 2011; Urgesi and Avenanti, 2011; Avenanti et al., 2012); and sensorimotor regions coupling action execution with perception (Chong et al. 2008; Etzel et al. 2008; Kilner et al. 2009; Oosterhof et al. 2010). The IFC, AIP and S1 are important sensorimotor regions of the AON and are consistently recruited not only when sensing or moving the body but also when perceiving the actions of others (Van Overwalle and Baetens 2009; Caspers et al. 2010; Grosbras et al., 2012). However, to date causative evidence that non-invasive stimulation of specific AON regions influences action perception is relatively scarce.

Few previous TMS studies have shown that online interference with IFC worsens: i) the ability to judge the weight of a box when seen lifted by a human agent (Pobric and Hamilton, 2006); ii) the visual discrimination of static images of actions with different kinematics (Urgesi et al., 2007); ii) the recognition of deceptive movements (Tidoni et al. unpublished observations). All these studies have shown impairments in the recognition of relatively simple actions like lifting a box or grasping an object. It is not yet clear whether similar disruption in action perception can be obtained with stimulation of other sensorimotor regions of the AON such as AIP or S1, nor whether stimulating these regions may impair perception of complex goal-oriented motor actions involving multiple sub-actions. Furthermore, it should be noted that virtual lesion approach indicates that IFC and possibly other regions of the AON may be critical for the recognition of others' actions but do not provide causative evidence on how others' motor acts are represented in the brain.

In the present study we used the TMSA paradigm to investigate the neural representation of observed complex goal-directed actions in the sensorimotor nodes of the AON. Participants were presented with adapting movies of an actor performing complex goal-directed actions on a tool by using a specific grip and were asked to categorize test pictures as showing similar or different action/grip relative to the adapting movie.

In the ‘arm action recognition task’, we found that applying TMS over S1 and IFC induced state-dependent effects on action recognition (see figure 4). TMS over S1 induced a selective decrease of IE index for pictures presenting the adapted action, indicating that stimulation of S1 improved the analysis of actions to which participants have been previously adapted. Similar greater performance for adapted relative to non-adapted actions was found with TMS over IFC, but not with sham stimulation or stimulation of AIP. Such pattern of findings suggests that TMS over S1 and IFC specifically enhanced performance of the neural subpopulations that respond to a specific invariant feature, i.e. the type of arm action, between the adapting stimulus and the test stimulus. The TMS-induced behavioral enhancement occurred when subjects had to attend to such invariant feature (i.e. in the Arm Action recognition task) and was absent when processing of the same feature was task-irrelevant (i.e. in the Grip recognition task). The state-dependency effects of TMS over S1 and IFC are well in keeping with the notion that repeated visual presentation of motor acts may induce action-specific adaptation phenomena in S1 (Dinstein et al., 2007) and IFC (Kilner et al., 2011). Our study significantly expands this notion by demonstrating behavioral consequences of such neural adaptation in the AON. Taken together these findings indicate that TMS over key nodes of the AON may behaviorally modulate the adapted relative to the non-adapted action features. Notably, both state-dependent effects of S1 and IFC stimulation were obtained for the adapted versus non-adapted action regardless of the type of grip used by the demonstrator to achieve it. These findings suggest that neurons in IFC and S1 are critically involved in the visual coding of goal-oriented actions at a relatively abstract level of representation in which low-level components of the action (i.e., the particular way the action is performed) are less reliable. These findings hint at a relatively abstract coding of goal-oriented motor acts in S1 which is independent from lower-level components such as the type of grip used to perform the action.

It should also be noted that the pattern of data found with stimulation of IFC suggest a decrease in performance for test pictures showing the non-adapted relative to the adapted action, which may be in part due to a summation of state-dependent and virtual lesion effects during stimulation of IFC. State-independent virtual lesions effects were clear in the Grip recognition task where subjects showed impaired performance when TMS was applied over the IFC relative to the other stimulation conditions (see figure 5). This result suggests that IFC is a critical node for the coding of observed grasping movements. However, the general impairment of behavioral performances after stimulation over this area indicates that both kinds of grip are coded within a unique neuronal population. This may explain why, in the 'Grip recognition task', TMS stimulation over the IFC results in a virtual lesion-like, state independent effect. In contrast, the distinct effect of stimulation over S1 and IFC over the recognition of arm actions suggest that the different types of arm actions – adapted and non-adapted – that are represented within segregated neuronal populations. Of note is that the complexity of the visual analysis (i.e., a global analysis for the 'Arm action recognition task versus a local analysis for the 'Grip recognition task') of the adapted stimuli and the task difficulty as well may have compromised the observation of state-dependent effects in the 'Grip recognition task'. Nevertheless, our study shows that a clear distinction in different neuronal populations tuned to a specific action-related feature is more likely to occur with the processing of the overarching action than with the grip.

The occurrence of this virtual lesion-like effect (also observed, to a lesser extent, in the IFC stimulation of the 'Arm action recognition task') is plausibly due to a jeopardized maintenance of action-related information in working memory relative to the adapted stimulus, rather than the deterioration of the stimulus perception *per se* (TMS is delivered at the onset of the test picture, when the information is not yet processed on the retina). Indeed, the task requires the subjects to maintain active the representation of the adapted stimulus during the recognition task (Candidi, 2008; Urgesi et al., 2007).

Conclusion

In summary, our results provide the first evidence that the IFC and the sensorimotor regions of the AON such as S1 are causatively involved during the perception of complex-goal directed actions. This recruitment is accounted for by the fact that TMS over S1 and IFC specifically enhanced performance of the neural subpopulations that respond to a specific invariant feature (i.e. the type of arm action) between the adapting stimulus and the test stimulus, and so independently of the behavioral sequence that composes the on-going action (i.e., the relationship between the type of grip used and the action goal). Furthermore, our study reveals that the facilitatory effect of TMS observed after adaptation to an invariant stimulus may partially overlap with inhibitory effect similar to virtual lesions techniques. This overlap could depend on the difficulty of the task at play, and possibly on the neural state of brain regions that are not currently targeted by the stimulation but that actively participate in the coding of stimulus features.

References

Adolphs, R., Damasio, H., Tranel, D., Cooper, G., & Damasio, A. R. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *20*(7), 2683-2690.

Akhtar, N., & Enns, J. T. (1989). Relations between covert orienting and filtering in the development of visual attention. *Journal of experimental child psychology*, *48*(2), 315-334.

Avenanti, A., Annella, L., Candidi, M., Urgesi, C., & Aglioti, S. M. (2012). Compensatory Plasticity in the Action Observation Network: Virtual Lesions of STS Enhance Anticipatory Simulation of Seen Actions. *Cerebral cortex (New York, N.Y.: 1991)*. doi:10.1093/cercor/bhs040

Avenanti, A., Bolognini, N., Maravita, A., & Aglioti, S. M. (2007). Somatic and motor components of action simulation. *Current biology: CB*, *17*(24), 2129-2135. doi:10.1016/j.cub.2007.11.045

Avenanti, A., Minio-Paluello, I., Bufalari, I., & Aglioti, S. M. (2009). The pain of a model in the personality of an onlooker: influence of state-reactivity and personality traits on embodied empathy for pain. *NeuroImage*, *44*(1), 275-283. doi:10.1016/j.neuroimage.2008.08.001

Avenanti, A., & Urgesi, C. (2011). Understanding « what » others do: mirror mechanisms play a crucial role in action perception. *Social cognitive and affective neuroscience*, *6*(3), 257-259. doi:10.1093/scan/nsr004

Avikainen, S., Forss, N., & Hari, R. (2002). Modulated activation of the human SI and SII cortices during observation of hand actions. *NeuroImage*, *15*(3), 640-646. doi:10.1006/nimg.2001.1029

Bertini, C., Leo, F., Avenanti, A., & Làdavas, E. (2010). Independent mechanisms for ventriloquism and multisensory integration as revealed by theta-burst stimulation. *The European journal of neuroscience*, *31*(10), 1791-1799. doi:10.1111/j.1460-9568.2010.07200.x

Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., & Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(11), 2553-2561. doi:10.1093/cercor/bhl161

Burton, M. P., McKeefry, D. J., Barrett, B. T., Vakrou, C., & Morland, A. B. (2009). Disruptions to human speed perception induced by motion adaptation and transcranial magnetic stimulation. *The European journal of neuroscience*, *30*(10), 1989-1998. doi:10.1111/j.1460-9568.2009.06993.x

Candidi, M., Urgesi, C., Ionta, S., & Aglioti, S. M. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Social neuroscience*, *3*(3-4), 388-400. doi:10.1080/17470910701676269

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010a). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148-1167. doi:10.1016/j.neuroimage.2009.12.112

Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(9), 2252-2258. doi:10.1093/cercor/bhp291

Cattaneo, Z., Rota, F., Walsh, V., Vecchi, T., & Silvanto, J. (2009). TMS-adaptation reveals abstract letter selectivity in the left posterior parietal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *19*(10), 2321-2325. doi:10.1093/cercor/bhn249

Cattaneo, Z., & Silvanto, J. (2008). Time course of the state-dependent effect of transcranial magnetic stimulation in the TMS-adaptation paradigm. *Neuroscience letters*, *443*(2), 82-85. doi:10.1016/j.neulet.2008.07.051

Chong, T. T.-J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current biology: CB*, *18*(20), 1576-1580. doi:10.1016/j.cub.2008.08.068

Christie, J., & Klein, R. (1995). Familiarity and attention: does what we know affect what we notice? *Memory & cognition*, *23*(5), 547-550.

Cohen Kadosh, R., Muggleton, N., Silvanto, J., & Walsh, V. (2010). Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(9), 2166-2171. doi:10.1093/cercor/bhp273

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *91*(1), 176-180.

Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and executed movements. *Journal of neurophysiology*, *98*(3), 1415-1427. doi:10.1152/jn.00238.2007

Etzel, J. A., Gazzola, V., & Keysers, C. (2008). Testing simulation theory with cross-modal multivariate classification of fMRI data. *PloS one*, *3*(11), e3690. doi:10.1371/journal.pone.0003690

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science (New York, N.Y.)*, *308*(5722), 662-667. doi:10.1126/science.1106138

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain: a journal of neurology*, *119 (Pt 2)*, 593-609.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral cortex (New York, N.Y.: 1991)*, *19*(6), 1239-1255. doi:10.1093/cercor/bhn181

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, *1156*, 97-117. doi:10.1111/j.1749-6632.2009.04425.x

Grosbras, M.-H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: a quantitative voxel-based meta-analysis. *Human brain mapping*, *33*(2), 431-454. doi:10.1002/hbm.21222

Kennett, S., Eimer, M., Spence, C., & Driver, J. (2001). Tactile-visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs. *Journal of cognitive neuroscience*, *13*(4), 462-478.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature reviews. Neuroscience*, *11*(6), 417-428. doi:10.1038/nrn2833

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in cognitive sciences*, *8*(11), 501-507. doi:10.1016/j.tics.2004.09.005

Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*(2), 335-346.

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *29*(32), 10153-10159. doi:10.1523/JNEUROSCI.2668-09.2009

Mayka, M. A., Corcos, D. M., Leurgans, S. E., & Vaillancourt, D. E. (2006). Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: a meta-analysis. *NeuroImage*, *31*(4), 1453-1474. doi:10.1016/j.neuroimage.2006.02.004

Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *Journal of neurophysiology*, *104*(2), 1077-1089. doi:10.1152/jn.00326.2010

Pobric, G., & Hamilton, A. F. de C. (2006). Action understanding requires the left inferior frontal cortex. *Current biology: CB*, *16*(5), 524-529. doi:10.1016/j.cub.2006.01.033

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rossi, S., Tecchio, F., Pasqualetti, P., Ulivelli, M., Pizzella, V., Romani, G. L., Passero, S., et al. (2002). Somatosensory processing during movement observation in humans. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *113*(1), 16-24.

Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., Dimitrijević, M. R., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalography and clinical neurophysiology*, *91*(2), 79-92.

Schwarzkopf, D. S., Silvanto, J., & Rees, G. (2011). Stochastic resonance effects reveal the neural mechanisms of transcranial magnetic stimulation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *31*(9), 3143-3147. doi:10.1523/JNEUROSCI.4863-10.2011

Serino, A., Canzoneri, E., & Avenanti, A. (2011). Fronto-parietal areas necessary for a multisensory representation of peripersonal space in humans: an rTMS study. *Journal of cognitive neuroscience*, *23*(10), 2956-2967. doi:10.1162/jocn_a_00006

Silvanto, J. (2008). A re-evaluation of blindsight and the role of striate cortex (V1) in visual awareness. *Neuropsychologia*, *46*(12), 2869-2871. doi:10.1016/j.neuropsychologia.2008.04.014

Silvanto, J., Muggleton, N. G., Cowey, A., & Walsh, V. (2007). Neural adaptation reveals state-dependent effects of transcranial magnetic stimulation. *The European journal of neuroscience*, *25*(6), 1874-1881. doi:10.1111/j.1460-9568.2007.05440.x

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in cognitive sciences*, *12*(12), 447-454. doi:10.1016/j.tics.2008.09.004

Silvanto, J., & Pascual-Leone, A. (2008). State-dependency of transcranial magnetic stimulation. *Brain topography*, *21*(1), 1-10. doi:10.1007/s10548-008-0067-0

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of experimental psychology. Human perception and performance*, *8*(6), 834-854.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature neuroscience*, *10*(1), 30-31. doi:10.1038/nn1815

Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., & Aglioti, S. M. (2010). Simulating the future of actions in the human corticospinal system. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(11), 2511-2521. doi:10.1093/cercor/bhp292

Valeriani, M., Betti, V., Le Pera, D., De Armas, L., Miliucci, R., Restuccia, D., Avenanti, A., et al. (2008). Seeing the pain of others while being in pain: a laser-evoked potentials study. *NeuroImage*, *40*(3), 1419-1428. doi:10.1016/j.neuroimage.2007.12.056

Valero-Cabré, A., & Pascual-Leone, A. (2005). Impact of TMS on the primary motor cortex and associated spinal systems. *IEEE engineering in medicine and biology magazine: the quarterly magazine of the Engineering in Medicine & Biology Society*, *24*(1), 29-35.

Valero-Cabré, A., Payne, B. R., & Pascual-Leone, A. (2007). Opposite impact on 14C-2-deoxyglucose brain metabolism following patterns of high and low frequency repetitive transcranial magnetic stimulation in the posterior parietal cortex. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *176*(4), 603-615. doi:10.1007/s00221-006-0639-8

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human brain mapping*, *30*(3), 829-858. doi:10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, *48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological bulletin*, *131*(3), 460-473. doi:10.1037/0033-2909.131.3.460

EXPERIMENTAL CONTRIBUTION

CHAPTER 2

The **Experimental Contribution Chapter 1** provides evidence that the primary somatosensory cortex (S1) plays a role during the observation of other people's actions. It is possible that S1feeds other regions of the Action Observation Network (AON) and especially the inferior frontal cortex (IFC) with which it is reciprocally connected, with sensorimotor and somatic information about the action that is currently observed.

Interestingly, sensorimotor and somatic information could also be used by an observer to estimate the biomechanical costs (muscular and articulatory) engaged in the execution of an action that is currently observed. In the **Experimental Contribution Chapter 2**, I hypothesized that the detection of some visual cues that enable an observer to make prediction about these costs activate prior information that biases the interpretation of other people's actions. In this study, these visual cues were provided by the object affordances. The detection of object affordances would activate prior information that provides a simple decision heuristic, recruited *by default* during the interpretation other agents' actions. Besides prior information conveyed object affordances, human observers take advantage of another kind of prior information to infer other people's intentions, that is, information extracted from the statistical regularities of past events (Griffiths et al., 2008). Indeed, human observers use the probability of occurence of another agent's intentions as a reliable source of information to infer the underlying intentions of upcoming actions (Chambon et al., 2011a, 2011b). Both these types of prior information may participate in action understanding. However, little is known about wheter, and how, these two classes of priors interact during the prediction of other people's intentions. Using methods of experimental psycholgy, this was what the study presented in the chapter 2 aimed to test.

Object Affordances Tune Observers' Prior Expectations About Tool-Use Behaviors

(accepted in PLoS One)

Pierre O. Jacquet[1,2], Valérian Chambon[3], Anna M. Borghi[1,4], and Alessia Tessari[1],

[1] Department of Psychology, University of Bologna, Bologna, Italy

[2] INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Center, ImpAct Team, Lyon, France

[3] Centre de Neuroscience Cognitive, CNRS UMR 5229, Université de Lyon, CNRS, France

[4] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

Abstract

Learning about the function and use of tools through observation requires the ability to exploit one's own knowledge derived from past experience. It also depends on the detection of low-level local cues that are rooted in the tool's perceptual properties. Best known as 'affordances', these cues generate biomechanical priors that constrain the number of possible motor acts that are likely to be performed on tools. The contribution of these biomechanical priors to the learning of tool-use behaviors is well supported. However, it is not yet clear if, and how, affordances interact with higher-order expectations that are generated from past experience – i.e. probabilistic exposure – to enable observational learning of tool use. To address this question we designed an action observation task in which participants were required to infer, under various conditions of visual uncertainty, the intentions of a demonstrator performing tool-use behaviors. Both the probability of observing the demonstrator achieving a particular tool function and the biomechanical optimality of the observed movement were varied. We demonstrate that biomechanical priors modulate the extent to which participants' predictions are influenced by probabilistically-induced prior expectations. Biomechanical and probabilistic priors have a cumulative effect when they 'converge' (in the case of a probabilistic bias assigned to optimal behaviors), or a mutually inhibitory effect when they actively 'diverge' (in the case of probabilistic bias assigned to suboptimal behaviors).

Key words: Action prediction, Affordances, Prior information, Observational learning, Tool use.

Introduction

Tool-use refers to a type of behavior that consists in manipulating external objects with the goal of altering the physical properties of another object, substance, surface, or medium, via a mechanical interaction, or that consists in mediating the flow of information between the tool user and the environment [1]. A growing amount of evidence suggests that the acquisition of knowledge about object use and function through observation is not the privilege of human subjects [2]. Yet, the richness and complexity of our technology suggests that we are particularly well adapted for such competence [3-6]. It has been argued that this competence arises from a set of interpretative and learning predispositions that allows human observers to i) decode kinematic information into the causal relationships between a behavioral sequence and its result [7], ii) interpret biological movements performed by others as 'rational' (i.e. assuming that the most optimal actions means are adopted to achieve a particular goal) [8], and iii) accumulate knowledge from past observations about an agent's intentions and behaviors, and use this database in order to predict future events [9-13]. Together, these mechanisms would enable human observers to derive knowledge about the possible uses and functions of a tool from observing goal-directed, intentional movements performed by an agent [14-16]. In this article we posit that these sophisticated learning skills could also benefit from simpler heuristics allocated to the detection of low-level, local sources of information, such as the manipulative properties of objects [17].

These properties, called 'affordances', are not intrinsic to objects but depend on their possible interactions with agents [18]. In its extended form [19] an affordance defines a relational property that emerges from matching the perceived physical features of an object (e.g. size, shape, texture, density) and the agent's biomechanical architecture, her goals, plans, values, beliefs, and past experiences. They are also described as dispositional states of the agent's nervous system [20]. Critically, affordances 'suggest' how one may interact with an object [21, 22]. For example, the size and shape of a softball mean that it fits into the human hand, and its density and texture make it perfect for throwing. We posit that object affordances contribute to delineating the number of potential motor acts that can be performed on a given object. They do this by generating effector-dependent, biomechanical priors which are in line with the agent's bodily architecture [17]. These priors then bias individuals to act on objects with the aim of biomechanical optimization. In both human and non-human primates, preferentially performed behaviors are generally those that minimize the muscular and/or articulator costs, given the object's affordances and the desired outcome [23-26].

Crucially, this minimization of costs also transfers to tool use learning. A prominent example is provided by our extensive technologies. Humans deliberately manufacture tools whose complex physical attributes offer naïve users affordances that enable the extraction of their functions at low cost [27-29]. Interestingly, the evolution of human technology might have increased the utility of simple heuristics such as affordance detection, in order to facilitate the highly demanding cognitive problem of tool use learning [28,30-32]. In our technological environments, the detection of affordances might thus play a crucial role in the acquisition of tool use skills through individual (i.e. trial-and-error learning) as well as social learning (i.e., learning from observing another agent's behaviors). Perceiving affordances may thus facilitate the extraction of functional features associated with an object manipulated by a third party [16]. For example, based on the amplitude of the observed agent's grip aperture and the orientation of her wrist, as well as on the size and texture of the object to be grasped, one may predict whether this object is meant to be lifted, pushed, or merely transported [11]. As suggested above, agents are expected to adopt tool-use behaviors that minimize biomechanical costs. Therefore, learning of a tool function through observation should be facilitated when a demonstrator uses a tool in a way that fit the observer's biomechanical expectations (behaviors that minimize the muscular and/or articulator costs), and should be jeopardized in the case where these expectations are patently violated (behaviors that increase the muscular and/or articulator costs).

Expert tool users, like tool learners, may also benefit from past experience in their daily interactions with objects [33]. It has been widely demonstrated that naïve human observers form knowledge (e.g. about tools and their potential use) by taking advantage of statistical regularities gathered from past observations [9-13]. The more times an individual associates a certain observed goal (e.g. the achieved tool function) with a certain observed action (e.g. the way of achieving the tool function), the more likely she is to expect that they will be seen together again [34]. These 'probabilistic' priors, acquired from past experiences, are crucial when the biomechanical information conveyed by tool affordances is too ambiguous or noisy to sufficiently constraint the range of candidate functions. Conversely, reference to biomechanical priors that are generated by tool affordances may be required when the use of the current tool cannot be based on previous experiences. Critically, both these classes of priors may be recruited when sensory information conveyed by movement kinematics is too incomplete to predict how an agent is most likely to behave. This occurs when many competing intentions are equally congruent with the not-yet completed behavior [11].

While the contribution of both these classes of priors to the individual-learning of tools' functions and use has long been demonstrated, it is not yet clear whether, and how, they may both combine to enable social learning of tool use (i.e., learning from observing another agent's behaviors). Here, we directly addressed this question in a task that required participants to predict, under various conditions of visual uncertainty, the intentions of a demonstrator who was using a multi-purpose tool. Affordance-related priors (termed 'biomechanical' priors) and priors acquired from past observations (termed 'probabilistic' priors) were manipulated by varying the biomechanical *optimality* of the tool behaviors and the *probability* (low versus high) of observing optimal versus suboptimal tool behavior.

We hypothesized that both biomechanical and probabilistic priors would have an effect on prediction. First, participants should be more accurate in predicting optimal than suboptimal behaviors (biomechanical bias). Second, participants should be more accurate in predicting behaviors that are most likely to occur throughout a specific experimental session (probabilistic bias). Third, we expected an interaction between these two classes of priors, whereby participants would preferentially respond towards the biased behaviors when the probabilistic bias is assigned to *optimal* behaviors. Finally, we expected this effect to vary as a function of the amount of visual uncertainty conveyed by the action being performed. Thus, the propensity to respond towards the biased behaviors should be strengthened as the amount of visual information shown in the action videos decreases.

Method

Participants

Twenty-four healthy volunteers (mean age=26.5, SD=4.40) took part in an action prediction task. All were right-handed, naïve to the purpose of the experiment, and reported normal or corrected-to-normal visual acuity. The experimental protocol was performed with approval of the University of Bologna - Department of Psychology - ethical committee and in accordance with the Declaration of Helsinki (2008) [35]. All participants gave their verbal and informed consent to participate in the study. Owing to the non-invasive, purely behavioral nature of our study (without any emotional stimuli), the University of Bologna - Department of Psychology - ethical committee considered verbal consent was appropriate and approved this consent procedure. Socio-demographic information (full name, age, sex, gender, handedness, education) has been collected for each subject on a separate sheet. The sheet contained an "Approve" box that was checked by the experimenter after the subject gave their verbal consent to participate.

Stimuli

Stimuli consisted in movies featuring a demonstrator acting on a two-purpose tool. The tool consisted of a movable handle screwed onto the lid of a box. The handle offered two distinct affordances enabling the demonstrator to grasp the object with a power or a precision grip (see fig.1). Using either grip, the demonstrator could achieve two intentions: *Opening the box by lifting the handle* (intention O); *Switching on the light by rotating the handle* (intention S) (see fig.1).

Two movie formats were displayed, both having a total duration of 2000msec (see fig.1): a *complete* format in which actions lasted until the achievement of the underlying intention (the grasp and the demonstrator's final intention were apparent); an *incomplete* format in which action course stopped 800msec after movement onset (only the grip was apparent but the demonstrator's final intention was not) while the last displayed frame was presented on the screen for the remaining 1200msec.

All movies were equalized for temporal homogeneity in such a way that the duration of the sub-steps of each action involved the same number of video frames (sub-step 1: static hand to physical contact with the tool=1000ms; sub-step 2: physical contact with the tool to action end-state=1000ms).

General Procedure

Participants sat in front of a monitor on which video clips that showed a male demonstrator acting on a tool were displayed (see fig.1). The entire experiment was composed of three distinct experimental sessions. In each session, participants had a different probability of observing the demonstrator achieving his intentions using an *optimal* (cost-free) or a *suboptimal* (high cost) behavioral strategy [33].

For each of the three sessions, 4 blocks of 24 complete action movies were interleaved with 4 blocks of 12 incomplete action movies. Crucially, the probabilistic bias was exclusively assigned during the complete action movie blocks, where participants could benefit from a high amount of visual information to identify the demonstrator's intentions. In contrast, in the incomplete action movies the amount of visual information was too low for the observer to unambiguously infer the demonstrator's intention. Thus, blocks of complete action movies were used to generate prior expectations in favour of either the optimal or the suboptimal behavioral strategy. These expectations were induced through biased probabilistic exposure. In contrast, blocks of incomplete movies were used to test the effect of each type of bias (probabilistic and biomechanical biases) on the participants' decisions when confronted with visually uncertain action scenes (see [11], for a similar procedure).

For each of the 144 action movies, participants were required to predict the demonstrator's intention by pressing, with their right index and middle fingers, one of two adjacent computer keys corresponding to the two possible intentions. The procedure used was a self-paced procedure: participants were instructed to make their response as soon as they though they had enough visual information to produce an accurate response. However, note that both complete and incomplete movies ran until completion independently of the subject's response.

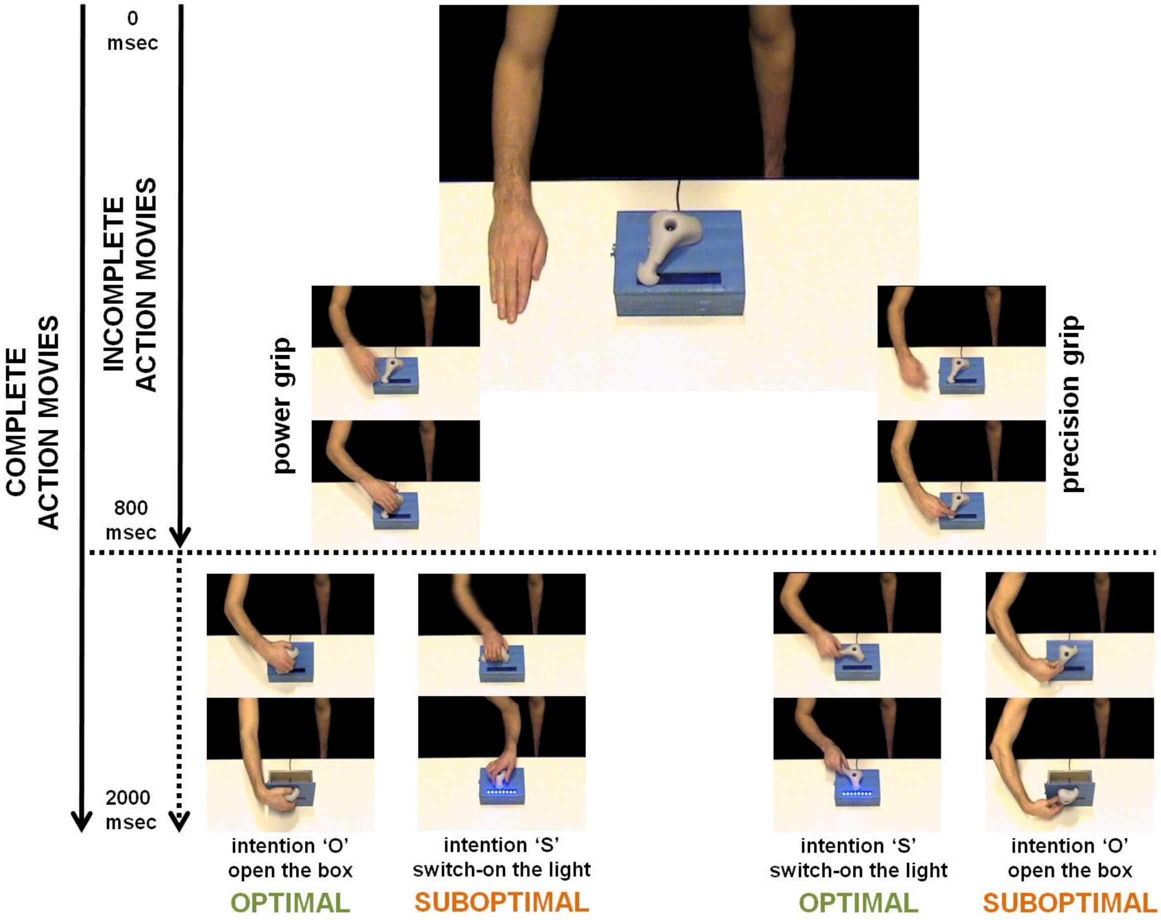


Figure 1: Examples of the four combinations 'grip × intention' that participants encountered during the experiment, and that lead to 'optimal' or 'suboptimal' behaviors. All combinations began with the demonstrator's static hand. The actor could then use either a 'power' or a 'precision' grip to achieve either the intention of *Opening the box* (O) or *Switching the lights on* (S). The combination between the kind of grip and the kind of final intention resulted in the complete action as being labeled biomechanically optimal (OPTIMAL) or suboptimal (SUBOPTIMAL). Whereas the complete action movies lasted until the achievement of the underlying intention for a total duration of 2000msec, the incomplete action movies stopped 800msec after the movement onset (when the demonstrator was about to grasp the tool) while the last displayed frame remained on the screen for a duration of 1200 msec, so that observers had information about the grip but no information (on that trial) about the demonstrator's intention.

Typical trial

All trials started with a white fixation-cross that appeared for 1000msec on a dark background. The fixation cross was immediately followed by either a complete or an incomplete action movie (see above for further details). After each decision, response time was displayed on the screen for 500msec. For those trials in which participants did not respond, or responded too late, ‘NO RESPONSE’ was displayed on the screen. The next trial started immediately after the 500msec visual feedback period. This feedback allowed us to avoid a ‘guessing bias’ that could occur during the presentation of complete action sequences, and that could hinder the integration of the probabilistic bias (see [11], for a similar procedure). The presentation of stimuli and recording of responses (correct/incorrect and response times) was synchronized using E-prime2 software (Psychology Software Tools, Inc, USA).

Biomechanical priors

The four possible action combinations (2 grips × 2 intentions) were divided into two types of behavioral category (optimal versus suboptimal) on the basis of their low or high biomechanical cost. This procedure allowed us to manipulate biomechanical priors emerging from perceived affordances (see fig.1):

i) **Optimal behaviors.** Using the power grip to achieve the intention of opening the box by lifting the handle was cost-free, as was using the precision grip to achieve the intention of switching the lights on by turning the handle. These two combinations were identified as *optimal* behaviors (low biomechanical cost).

ii) **Suboptimal behaviors.** The precision grip increased the cost of achieving the intention of opening the box, whereas the power grip increased the cost of achieving the intention of switching on the lights. These two combinations were identified as *suboptimal* behaviors (high biomechanical cost).

The biomechanical cost of action movies were pre-tested on 10 naïve individuals. They were asked to estimate the muscular and/or articulator cost of each perceived movement on a 5-point Likert scale (ranging from 0 = null cost to 5 = very high cost). As expected, optimal behaviors (precision grip/switching-on the lights and power grip/opening the box, mean score = 1.01) were estimated as significantly less costly than suboptimal ones (precision grip/opening the box and power grip/switching-on the lights, mean score = 3.13) (two-tailed t-test for paired data: t = -20.87, p < .0001). It is of note that the intentions achieved with a precision grip were rated as less costly than those achieved with a power grip for both optimal (precision grip/switching-on the lights, mean score = 0.55, versus power grip/opening the box, mean score = 1.47; two-tailed t-test for paired data: t = -54.83, p < .0001) and suboptimal behaviors (precision grip/opening the box, mean score = 2.90, versus power grip/switching-on the lights, mean score = 3.37; two-tailed t-test for paired data: t = -30.82, p < .0001).

Probabilistic priors

Unbeknownst to the participants, the probability of observing the demonstrator using an optimal or a suboptimal behavioral strategy was varied within the three distinct experimental sessions (‘*baseline*’, ‘*convergent bias*’, ‘*divergent bias*’ – see below). Varying the probability distributions of each possible strategy allowed us to manipulate each participant’s probabilistic priors, that is, prior expectations they could form about the behavioral strategy being favored by the demonstrator to achieve the tool’s functions. After each participant performed the task, we controlled for the extent to which she/he was aware of the induced bias. As expected, none of the subjects spontaneously reported that one type of action was more likely observed than another.

i) Baseline session: no probabilistic bias. In the first session, participants had an equal probability of observing the demonstrator achieving his intention by performing an optimal or a suboptimal behavior.

ii) 'Convergent bias' session: probabilistic bias towards optimal behaviors. In this session participants were biased towards 'optimal' behaviors to the detriment of 'suboptimal' behaviors. In 80% of the 'box opening' trials the demonstrator opened the box using a power grip, and in 80% of the 'light switching' trials he switched on the lights using a precision grip. Here, behaviors that were preferentially used by the demonstrator *converged* towards the participant's biomechanical priors.

iii) 'Divergent bias' session: probabilistic bias towards suboptimal behaviors. In this session participants were biased towards 'suboptimal' behaviors to the detriment of 'optimal' behaviors. In 80% of the 'box opening' trials the demonstrator opened the box using a precision grip, and in 80% of the 'light switching' trials he switched on the lights using a power grip. Here, the behaviors that were preferentially used by the demonstrator *diverged* from the participant's biomechanical priors.

All participants began the experiment with the baseline session. The order of the two bias sessions (convergent and divergent) was counterbalanced across participants.

Training phase

Prior to the experiment participants were familiarised with the task. The training consisted of an unbiased complete action movie block followed by an incomplete action movie block.

Data analysis

We analysed the percentage of correct responses (hits) and response times (RTs) collected for both complete and incomplete action movies. Responses for incomplete actions were encoded as correct if the predicted intentions conformed to those that the demonstrator actually achieved in their complete format. Participants who responded too early on more than 10 percent of the complete action movies were discarded from further analyses (responses were considered as too early when they occurred between 0 and 1000msec after movie onset, making accurate predictions impossible). Using this criterion, two subjects were excluded.

All statistical analyses were performed separately for complete and incomplete action movies. The magnitude of the probabilistic bias and its interaction with biomechanical expectations was investigated by comparing performance during the baseline session with that during the two biased sessions. The hit rates and RTs were then analysed using a 2 × 2 × 3 repeated-measures ANOVAs. The first two-level factor was the 'type of behavior' (optimal versus suboptimal behaviors), the second two-level factor was the 'type of grip' (power versus precision grip), and the third, three-level factor was the 'probabilistic bias' (baseline versus convergent bias versus divergent bias). Post-hoc Fisher tests were used to compare performance between conditions.

We further investigated the learning dynamics internal to each session by comparing data (hits and RTs) collected during the first (time-step 1) and the second half (time-step 2) of each session. Thus, for each session, the hits rates and RTs were analysed using 2 × 2 × 2 repeated-measures ANOVAs with ‘time-step’ (time-step 1 versus time-step 2), ‘type of behavior’ (optimal versus suboptimal behaviors), and ‘type of grip’ (power versus precision grip) as two-level factors. Post-hoc Fisher tests were used to compare performance between conditions.

For all analyses, $p < .05$ was taken as the criterion for significance and eta squared (ή) was used as a measure of effect size. Statistical analyses were performed using *Statistica 9* (www.statsoft.com).

Results

Overall performance

Complete action movies (Hits and RTs)

The 2 (type of behavior) × 2 (type of grip) × 3 (probabilistic bias) repeated-measures ANOVAs revealed a main effect of the ‘type of behavior’ on both hits ($F_{1,21} = 18.08$, $p < .001$, ή = .46) and RTs ($F_{1.21} = 93.43$, $p < .0001$, ή = .82). Participants were more accurate and faster at predicting optimal than suboptimal behaviors (hits: 88% vs. 81%; RTs: 1382msec vs. 1444msec). The main effect of the ‘probabilistic bias’ was also significant on both hits ($F_{2,42} = 6.5$, $p < .01$, ή = .24) and RTs ($F_{2.42} = 22.18$, $p < .0001$, ή = .51). In the divergent bias session, participants made more accurate predictions compared to the baseline (hits: 88% vs. 84%, $p < .05$) and the convergent bias sessions (hits: 88% vs. 82%, $p < .001$). However, when compared to baseline, RTs were faster in both the convergent (1368msec vs. 1452msec, $p < .0001$) and the divergent bias sessions (1420msec vs. 1452msec, $p < .05$). It is of note that a difference occurred also between the two bias sessions, with faster RTs in the convergent bias session (1368msec vs. 1420msec, $p < .001$). Finally, a main effect of the 'type of grip' was found on hits only ($F_{1,21} = 23.27$, $p < .0001$, ή = .53), with participants being overall more accurate at predicting behaviors that were performed with a precision than a power grip (88% vs. 81%).

The two-way interaction 'type of behavior' × 'probabilistic bias' was significant for both hits ($F_{2.42} = 19.76$, $p < .0001$, ή = .48) and RTs ($F_{2.42} = 31.69$, $p < .0001$, ή = .60) (see fig.2a,b). Post-hoc comparisons (LSD Fisher tests) indicated that during the baseline session – where both types of behaviors were equally probable – participants were more accurate (87.5% vs. 80%, $p < .01$) and faster (1411msec vs. 1492msec, $p < .0001$) at predicting optimal compared to suboptimal behaviors. A similar pattern was observed in the convergent bias session. Participants were more accurate (91% vs. 72%, $p < .0001$) and faster (1308msec vs. 1427msec, $p < .0001$) at predicting the optimal behaviors when these behaviors were more frequently shown than the suboptimal ones. In the divergent bias session, no differences were found between the optimal and suboptimal behaviors, despite the fact that the latter were more frequently shown than the former (hits = 85% vs. 90%, $p > .05$; RTs = 1427msec vs. 1414msec, $p > .05$). Thus, increasing the probability of observing suboptimal behaviors did not significantly increase the number of correct responses for these behaviors compared to the optimal ones.

Interestingly, the interaction effect between the optimality of the behavior and the probabilistic bias was further modulated by the type of grip used, as revealed by a significant three-way interaction between all three factors for hits ($F_{2.42} = 9.49$, $p < .001$, ή = .31). In the baseline session, the preference for optimal over suboptimal behaviors was observed for power grip only (post hoc test comparing optimal/power grip vs. suboptimal/power grip: p <

.0001; post-hoc test comparing optimal/precision vs. suboptimal/precision grip: $p > .05$). In the convergent bias session, participants were impaired at predicting suboptimal over optimal behaviors irrespective of the type of grip used. In the divergent session, no difference between optimal and suboptimal behaviors was observed, irrespective of the type of grip used.

Incomplete action movies (Hits and RTs)

The 2 (type of behavior) × 2 (type of grip) × 3 (probabilistic bias) repeated-measures ANOVAs revealed a main effect of the 'type of behavior' on both hits ($F_{1.21} = 17.19$, $p < .001$, ή = .45) and RTs ($F_{1.21} = 6.97$, $p = .01$, ή = .25); participants were more accurate and faster at predicting optimal than suboptimal behaviors (hits: 58% vs. 42%; RTs: 1176msec vs. 1215msec). This preference for optimal behaviors significantly differed from chance (t-test for single mean, $t > 4.40$, $p < .001$).The main effect of the 'probabilistic bias' was significant only for RTs ($F_{2,42} = 5.75$, $p < .01$, ή = .21). This indicated that, compared to the incomplete movie blocks of the baseline session, participants make faster predictions in the incomplete movie blocks of the convergent bias (1156msec vs. 1235msec, $p < .01$). Note that they also tended to make faster predictions in the incomplete movies of the divergent bias session (1194msec vs. 1235msec, $p = .08$). The main effect of 'type of grip' was not significant (hits and RTs: all $F > .33$, all $p > .48$).

The two-way interaction 'type of behavior' × 'probabilistic bias' was significant for both hits ($F_{2,42} = 9.84$, $p < .001$, ή = .32) and RTs ($F_{2,42} = 3.34$, $p < .05$, ή = .14) (see fig.2c,d). As for the complete movie blocks, post-hoc comparisons (LSD Fisher tests) indicated that, in the baseline session, participants were more accurate at predicting optimal than suboptimal behaviors (59% vs. 35%, $p < .001$). This preference for optimal behaviors significantly differed from chance (t-test for single mean, t > 3.32, p < .01).They were also more accurate (66% vs. 36%, p < .0001) and faster (116msec vs. 1197msec, p < .001) at predicting optimal than suboptimal behaviors in the incomplete action movie blocks of the convergent bias session. Again, the preference for optimal behaviors was significantly different from chance level (t-test for single mean, t > 4.75, p < .001). However, in the incomplete action movie blocks of the divergent bias session, we did not find any differences between the optimal and the suboptimal behaviors, although the latter were most likely observed than the former in the complete movie blocks that preceded (hits = 49% vs. 55%, p > .05; RTs = 1187msec vs. 1202msec, p > .05). Note that performances for both optimal (t-test for single mean, t < -0.17, p > .05) and suboptimal behaviors (t-test for single mean, t > 1.46, p = .15) did not significantly differ from chance.

Finally, the interaction effect between the ‘type of behavior’ performed (optimal vs. suboptimal) and the ‘probabilistic bias’ (baseline vs. convergent vs. divergent) was modulated by the type of grip (power vs. precision) used by the demonstrator ($F_{2,42} = 3.37$, p < .05, ή = .14). In the incomplete action movie blocks of the baseline and convergent bias sessions, the difference between optimal and suboptimal behaviors was observed independently of the type of grip used. In the incomplete action movie blocks of the divergent bias session, a difference between optimal and suboptimal behaviors was observed only when both of them were achieved by a precision grip (optimal/precision = 47% vs. suboptimal/precision = 59%). Note that the proportion of correct predictions for suboptimal behaviors achieved with a precision grip differed from chance (t-test for single mean, t > 2.38, p < .05).

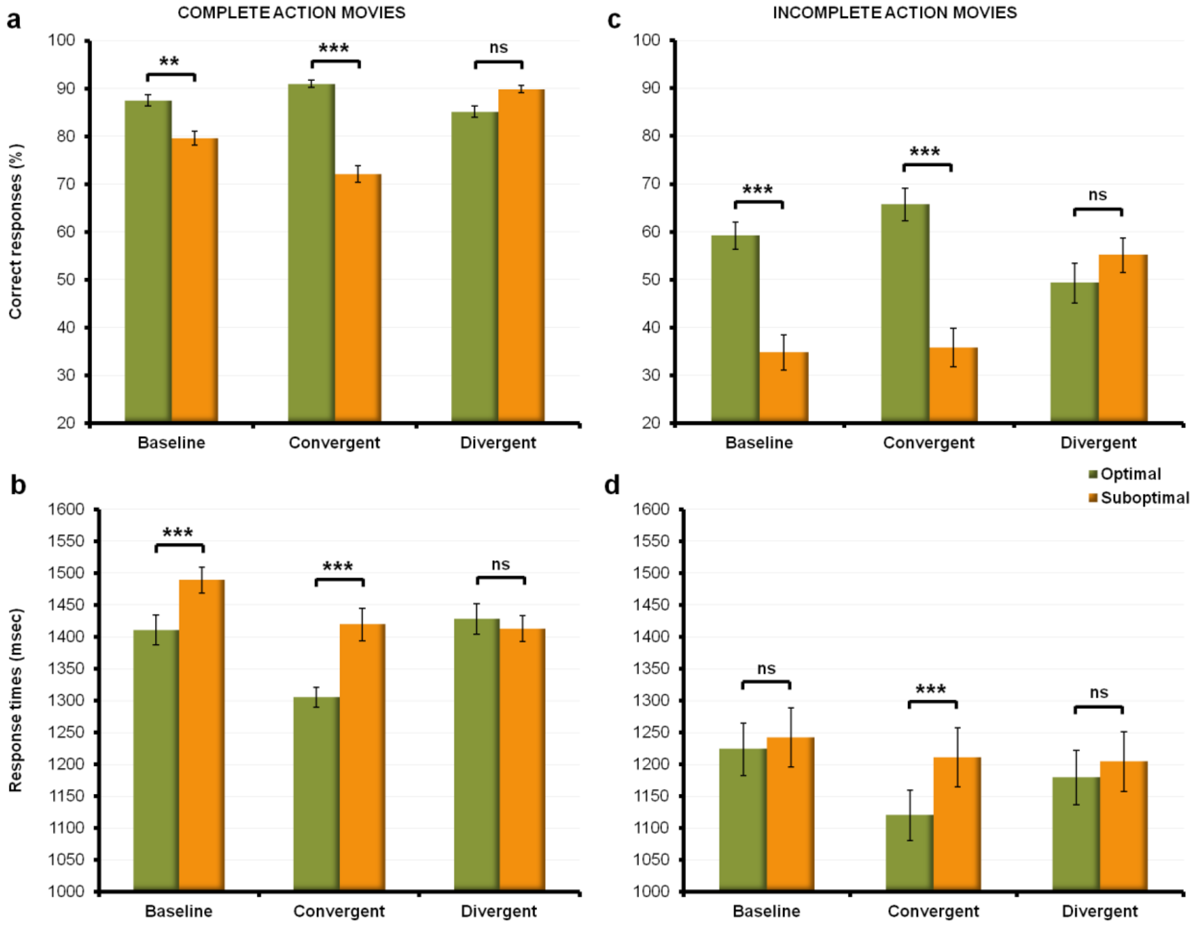


Figure 2: Overall performances. a) and c) represent the mean percentages of correct responses collected during complete and incomplete action movies for all three sessions. b) and d) represent the mean response times collected during complete and incomplete action movies for all three sessions. The green columns refer to the mean percentages of correct predictions for observed 'optimal' behaviors (pooled across 'power' and 'precision' grip). The orange columns refer to the mean percentages of correct predictions for observed 'suboptimal' behaviors (pooled across 'power' and 'precision' grip). Error bars denote the standard error of the mean.

Overall performance: preliminary discussion (fig.2)

Results for the complete action movies demonstrate that, compared to baseline, the probabilistic bias significantly improved participants' performance – as also indicated by faster reaction times in the two bias sessions. Note that the rate of correct responses was overall higher in the divergent session. This is easily explained by the fact that, in the convergent session, the probabilistic bias assigned to optimal behaviors concomitantly increased the errors rate for unbiased (i.e., suboptimal) behaviors. In contrast, the probabilistic bias assigned to suboptimal behaviors did not alter the participants' ability to accurately predict the unbiased (i.e., optimal) behaviors. Thus, the higher the probability that a behavior occured, the better and faster it was predicted, irrespective of its type (optimal or suboptimal). These results indicate that, as expected, participants were successful in integrating the probability distributions of both convergent and divergent bias sessions.

The second set of results shows that the biomechanical constraints generated by the detection of tool affordances play a major role in participants' predictions: participants were more accurate and faster at predicting behaviors that minimized biomechanical costs, irrespective of probabilities. Thus, in both the complete and incomplete action movies of the baseline session (i.e. a session in which the demonstrator equally selected between the two available behavioral strategies), participants preferentially chose intentions achieved by optimal behaviors rather than suboptimal behaviors (see fig.2a,b,c,d). This result demonstrates that when participants cannot rely on past observations (i.e., on probability) to decide how an observed agent is most likely to behave, they tend to rely on their biomechanical priors *by default*. That is, they assume that the observed agent behaves 'rationally', i.e., that he favors strategies which minimize biomechanical costs.

The third set of results concerns the interaction between the two kinds of priors (biomechanical and probabilistic) (fig.2a,b,c,d). We found that both the magnitude and dynamics of the probabilistic bias differed as a function of the type of behavior, with participants' biomechanical expectations overriding the effect of the probabilistic bias. Thus, in the convergent bias session (probabilistic bias assigned to optimal behaviors) performance decreased for the suboptimal behaviors, and was facilitated for the optimal behaviors, as expected. This pattern of performance – observed in both the incomplete movie and complete movie blocks – suggests that it is costly for participants to inhibit a response that fits with their biomechanical expectations, even though a high amount of visual information is available. However, in the divergent bias session (probabilistic bias assigned to suboptimal behaviors), no significant differences were found between the two alternatives: participants did not preferentially choose the suboptimal behavior over the optimal one, although the former was more likely to be performed than the latter. This pattern suggests that participants actively integrated both types of priors, by combining their respective effects. Thus, when probabilistic and biomechanical priors diverged, the overall effect tended to sum to zero, resulting in performances that did not significantly differ from chance for both optimal and suboptimal behaviors.

Finally, we found that the type of grip used by the demonstrator had an effect on the participants' predictions when i) the probability of each competing intention was equal (baseline session), and ii) when the intention that was eventually achieved was fully visible (complete movies). This finding can be accounted for by a *facilitatory effect* of the precision grip. Although suboptimal behaviors that were achieved with a precision grip were estimated as suboptimal, they were nevertheless estimated as less constraining than those performed with a power grip. Interestingly, this facilitatory effect was easily overcome by the probabilistic bias, since it disappeared in both the convergent and divergent bias sessions. It is of note that this tendency to over-estimate the optimality of precision grips may be due to the biomechanical characteristics of the effector itself. Indeed, performing prehension movements with either a power grip or a precision grip differentially affects the synergies of arm segments. While the achievement of a power grip exerts constraints on many degrees of freedom of the arm (i.e. the wrist, elbow and shoulder) [36], the precision grip offers more flexible solutions [37], independently of the overall cost of the final action (e.g. opening the box with a precision grip).

Learning dynamics

Complete action movies (Hits and RTs)

i) *Baseline session.* The 2 × 2 × 2 repeated-measures ANOVA performed on ‘time-step’ (time-step 1 vs. time-step 2),‘type of behavior’ (optimal vs. suboptimal) and ‘type of grip’ (power vs. precision grip) revealed a main effect of the ‘type of behavior’ for both hits ($F_{1,21}$ = 11.57, p < .01, ή = .36) and RTs ($F_{1,21}$ = 47.7, p < .0001, ή = .69), with optimal behaviors being overall faster (1411msec vs. 1493msec) and more accurately predicted (88% vs. 80%) than suboptimal ones. A main effect of ‘type of grip’ was also found on hits only ($F_{1,21}$ = 9.48, p < .01, ή = .31), with behaviors achieved using a precision grip being overall more accurately predicted than those achieved using a power grip (87% vs. 80%). The two-way interaction ‘time-step’ × ‘type of behavior’ was significant for hits ($F_{1,21}$ = 4.91; p < .05, ή = .19) (see fig.3a). Post-hoc comparison tests (LSD Fisher tests) showed that the difference between the percentage of hits observed at time-step 1 for the optimal and the suboptimal behaviors (90% vs. 78%; post-hoc test: p < .0001) was no longer significant at time-step 2 (85% vs. 82%; post-hoc test: p > .05). Neither the main effect of ‘time-step’, nor the two-way interaction ‘time-step’ × ‘type of grip’, nor the three-way interaction was significant (hits and RTs: all F < 2.93, all p > .10).

ii) Convergent bias session. The same 2 × 2 × 2 repeated-measures ANOVA performed on complete movie blocks of the convergent bias session revealed main effects of ‘time-step’ (hits: $F_{1,21} = 9.80$; p < .01, ή = .32; RTs: $F_{1,21} = 6.87$; p < .05, ή = .25) and ‘type of behavior’ (hits: $F_{1,21} = 34.09$; p < .0001, ή = .62; RTs: $F_{1,21} = 43.61$; p < .0001, ή = .67) on both hits and RTs. Participants were more accurate but slower at predicting the demonstrator’s intention at time-step 1 than at time-step 2 (hits = 85% vs. 78%, p < .01); RTs = 1386msec vs. 1337msec, p < .05). Overall, they were more accurate and faster at predicting likely optimal than unlikely suboptimal behaviors (hits = 91% vs. 73%, p < .0001; RTs = 1307msec vs. 1416msec, p < .0001). A main effect of the ‘type of grip’ was also shown on hits only ($F_{1,21} = 17.26$; p < .001, ή = .45), revealing that participants more accurately predicted behaviors performed with a precision than a power grip (87%vs. 77%, p < .001), independently of their optimality and of the time-step. Furthermore, the two-way interaction ‘time-step’ × ‘type of behavior’ was significant for hits ($F_{1,21} = 9.07$; p < .01, ή = .30) (see fig.3a). Post-hoc analyses (LSD Fisher tests) showed that throughout the session, participants were overall more accurate at predicting the optimal than the suboptimal behaviors, and that this advantage for optimal behaviors increased over time (time-step 1 = 91% vs. 79%, p < .001; time-step 2 = 91% vs. 66%, p < .0001). The two-way interaction between ‘time-step’ × ‘type of grip’ as well as the three-way interaction were not significant (hits and RTs: all F < 1.60, all p > .22).

iii) Divergent bias session. The same 2 × 2 × 2 repeated-measures ANOVA performed on complete movie blocks of the divergent bias session showed a main effect of ‘time-step’ ($F_{1,21} = 5.04$.; p < .05, ή = .19), with better performance at time-step 1 than at time-step 2 (90% vs. 85%). A main effect of the ‘type of grip’ was also found on hits ($F_{1,21} = 6.99$.; p < .05, ή = .25), with better performance for behaviors performed with a precision than a power grip (90%

vs. 84%), irrespective of their optimality. The interaction between the ‘time-step’ and ‘type of behavior’ factors was significant for hits only ($F_{1,21} = 6.85.$; $p < .05$, ή = .25) (see fig.3a). In the first half of the session participants performed equally well (post-hoc test: $p > .05$) for the likely suboptimal (time-step 1 = 89%) and the unlikely optimal behaviors (time-step 1 = 91%). In the second half, however, they were more accurate at predicting the suboptimal behaviors (time-step 2 = 91% vs 79%; $p < .01$). This was associated with decreased performance for the unlikely optimal behaviors throughout the session (time-step 1 = 90% vs. time-step 2 = 79%). The main effect of ‘type of behavior’, the ‘time-step’ × ‘type of grip’ interaction, and the three-way interaction were not significant (hits and RTs: all $F < 3.83$, all $p > .06$).

Incomplete action movies (Hits and RTs)

i) *Baseline session.* The 2 × 2 × 2 repeated-measures ANOVA performed on ‘time-step’ (time-step 1 vs. time-step 2),‘type of behavior’ (optimal vs. suboptimal) and ‘type of grip’ (power vs. precision grip) showed a main effect of the ‘type of behavior’ on hits only ($F_{1,21} = 17.96$, $p < .001$, ή = .46). In the incomplete movie blocks of the baseline session, participants were more accurate at predicting optimal (59%) than suboptimal (35%) behaviors, independently of the time-step. Neither the main effects of ‘time-step’ or ‘type of grip’, nor the two-way interactions ‘time-step’ × ‘type of grip’ and ‘time-step’ × ‘type of behavior’ (see fig.3b), nor the three-way interaction were significant (hits and RTs: all $F < 1.21$, all $p > .28$).

ii) *Convergent bias session.* The same 2 × 2 × 2 repeated-measures ANOVA performed on incomplete movie blocks of the convergent bias session revealed a main effect of ‘time-step’ on RTs only ($F_{1,21} = 9.53$; $p < .01$, ή = .31). Overall, participants responded slower at time-step 1 (1178msec) than at time-step 2 (1141msec). A main effect of the ‘type of behavior' was present for both RTs ($F_{1,21}$= 14.11; p < .01, ή = .40) and hits ($F_{1,21}$ = 21.17; p < .001, ή = .50), with participants being more accurate (66% vs. 36%) and faster (1116msec vs. 1203msec) at predicting optimal than suboptimal behaviors. The main effect of the 'type of grip', the 'time-step' × 'type of grip' and 'time-step' × 'type of behavior' interactions (see fig.3b), and the three-way interaction were not significant (hits and RTs: all F < 3.77, all p > .07).

iii) Divergent bias session. The same 2 × 2 × 2 repeated-measures ANOVA performed on incomplete movie blocks of the divergent bias session showed a significant interaction between the 'time-step' and 'type of behavior' on hits only ($F_{1,21}$ = 8.39; p < .01, ή = .27) (see fig.3b). Post-hoc tests (LSD Fisher tests) demonstrated that in the first half of the incomplete movie blocks, rates of correct predictions for the optimal and the suboptimal behaviors did not differ (time-step 1 = 54% vs. 54%; p > .05). However, a difference occurred in the second half of the incomplete movie blocks, with suboptimal behaviors being more accurately predicted than optimal ones (time-step 2: optimal = 44% vs. suboptimal = 57%; p < .001). Of note is the fact that this effect was due to the rate of correct predictions for the optimal behaviors decreasing over the session (time-step 1 = 54% vs. time-step 2 = 44%; p < .01). However, neither the performance for suboptimal behaviors (t-test for single mean, t < 1.47, p = .15) nor the performance for optimal behaviors (t-test for single mean, t < -1.32, p = .19) significantly differed from chance level. No significant main effects were revealed (hits and RTs: all F < 1.87, all p > .19). Neither the 'time-step' × 'type of grip' interaction was significant (hits and RTs: all F < .74, all p > .40).

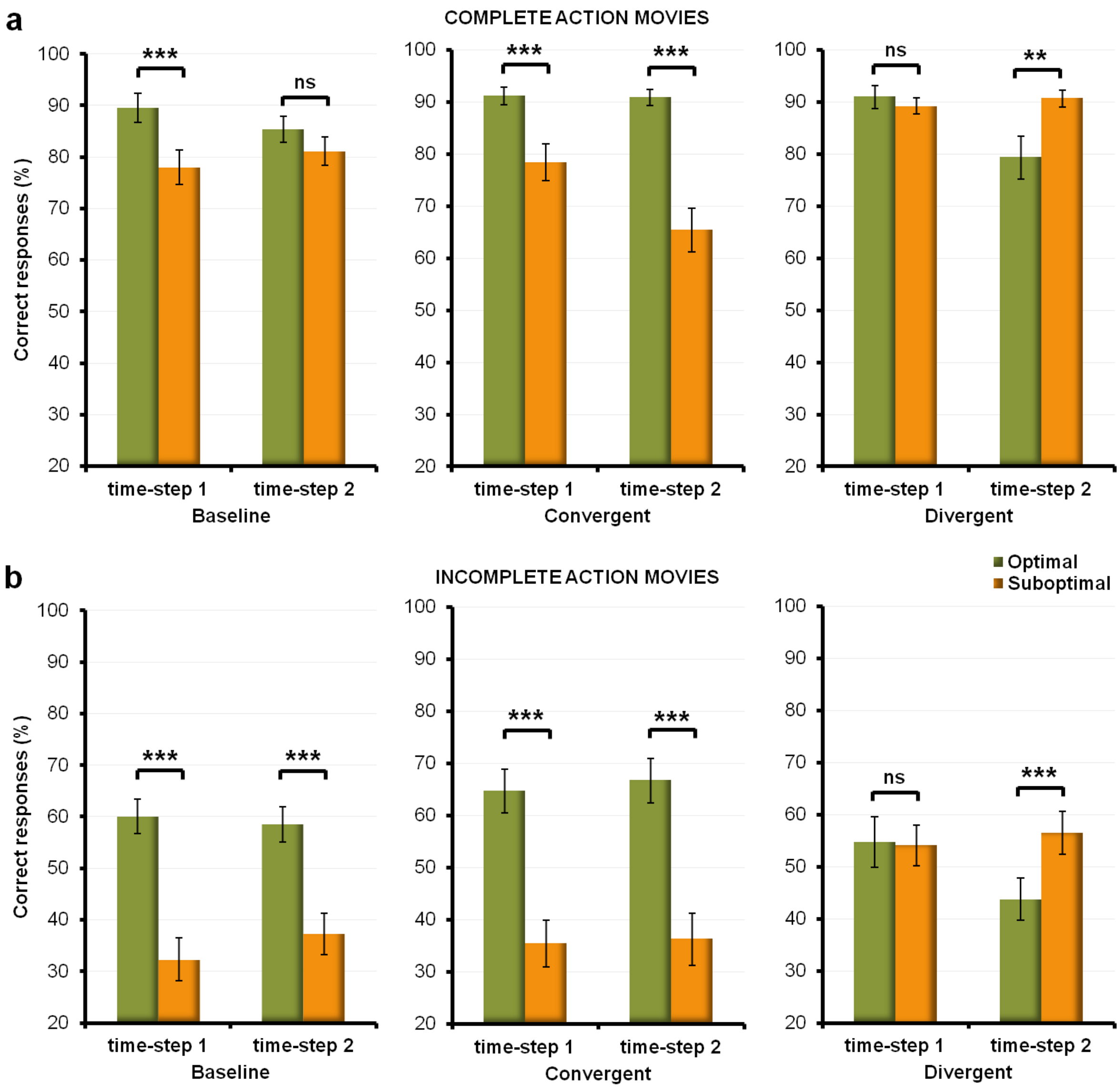


Figure 3: **Learning dynamics**. a) and b) represent the mean percentages of correct responses collected during complete and incomplete action movies for all three sessions. The green columns refer to the mean percentages of correct predictions for 'optimal' behaviors (pooled across 'power' and 'precision' grip). The orange columns refer to the mean percentages of correct predictions for 'suboptimal' behaviors (pooled across 'power' and 'precision' grip). Error bars denote the standard error of mean.

Learning dynamics: preliminary discussion (fig.3)

In both the baseline and the convergent bias session, analyzing the evolution of response patterns over time (from time-step 1 to time-step 2) revealed an early preference for the optimal behaviors (see fig.3a,b). This preference was already present in the first half of the baseline session and did not vary further with increasing probabilities. Interestingly, this preference for behaviors that minimized biomechanical costs seemed impervious to their probabilistic likelihood sampled from past observations. This suggests that biomechanical priors might short-circuit probabilistic sampling, and might interfere with decisions based on the extraction of statistical regularities.

In the divergent bias session (suboptimal bias), the evolution of the response pattern from time-step 1 to time-step 2 suggests that the absence of a difference between performance for optimal and suboptimal behaviors – although the latter were more frequently shown – could be primarily due to participants' initial preferences for optimal behaviors (see fig.3a,b). This preference progressively decreased over time as the probability of observing suboptimal behaviors concomitantly increased. However, overall, this increase was not sufficient to compensate for the participants' initial lack of preference toward suboptimal behaviors.

Finally, it is noteworthy that the number of responses toward optimal versus suboptimal behaviors was overall greater in the incomplete, relative to the complete, action movies in both the baseline and the convergent bias sessions. This difference may account for the fact that the rate of hits for both the optimal and suboptimal behaviors was very high in the complete movie blocks. Therefore, the number of responses for optimal behaviors, and hence the difference between the two types of behavior, could not further increase due to a 'ceiling' effect. Alternatively, this difference may be accounted for by the fact that, in conditions of visual uncertainty, individuals tended to favor responses that were consistent with their prior expectations. Interestingly, this assumption is consistent with the finding that one's priors (here, an intrinsic preference for optimal behaviors) are primarily used to complement sensory uncertainty in order to allow decisions, and thus actions, to be made even in cases of noisy signals or sparse data [11,16].

Discussion

The aim of this study was to test how the biomechanical expectations conveyed by tool affordances interact with prior knowledge about tool function and use, and whether this interaction influences predictions about a demonstrator's intentions when using tools. Here, we provide the first evidence that low-level local cues such as object affordances influence the learning and prediction of tool-use behaviors. We demonstrate that biomechanical priors modulate the extent to which participants' predictions are influenced by probabilistically-induced prior expectations (see fig.2). In particular, we found that when the demonstrator's behavior satisfied both the participants' biomechanical and probabilistic priors, the learning cost decreased, as participants efficiently combined both types of priors to make their predictions. Conversely, when the demonstrator's behavior violated the biomechanical but not the probabilistic priors, the learning cost increased, as participants had to deal with two conflicting sources of prior information.

Specifically, the dynamics of the integration of these probabilistic expectations was strongly dependent on the biomechanical optimality of the observed behaviors (see fig.3).

When the probabilistic bias favored suboptimal behaviors, participants needed a greater number of observations to neutralize a preference for optimal behaviors, as well as to derive and use probabilistic information to predict suboptimal behaviors. Furthermore, performance during both the baseline and the convergent bias sessions showed that participants exhibited an initial preference for optimal behaviors that was sustained throughout the session, and did not vary with changes in probabilistic bias. Interestingly, this initial preference was even stronger in the interrupted sequences, where subjects had little information about the demonstrator's intention. The strong influence of biomechanical priors in these sequences suggests that these priors might be primarily used in the case of noisy signals or sparse data. As such, they may be specifically suited to reduce the intrinsic uncertainty of goal-directed behaviors [16]. In sum, biomechanical priors provided by the tool's affordances acted as an inductive bias [13], complementing the available perceptual information when this information did not sufficiently constrain the number of potential solutions (e.g. 'opening a box' versus 'switching the lights on').

Together, these findings complement recent results published by Chambon and co-workers [11]. In their study, participants were requested to infer the intentions of a demonstrator who performed various actions on meaningless objects. The authors showed that as the amount of visual information conveyed by movement kinematics progressively decreased, participants responded more frequently toward the intentions that had the highest probability of occurring. Chambon et al.'s findings are consistent with a Bayesian estimation scheme: the less information one has about the action scene, the greater the weight of one's priors in the decision. Put another way: the higher the sensory uncertainty, the more the probabilistic bias is used to 'resolve', or 'complement', this sensory uncertainty. Our findings suggest that the effect of priors gathered from probabilistic sampling of past observations also depends on whether or not the visual information conveyed by the movement's kinematics meets the expectations that are induced by an object's affordances.

Even though visual information did not meet these expectations, participants tended to assume the demonstrator to behave in an optimal way. In other words, they expected the demonstrator to act as a 'rational' agent – i.e., an agent who adopts the most optimal (i.e., least costly) action means to achieve his goal given the constraints of the current situation. This echoes recent evidence showing that humans, even at a very early age, consider their conspecifics to be rational agents [8,38,39]. Thus, children may posit states of the world occasionally counterfactual to the perceptual evidence (such as the presence of occluded physical objects) but consistent with a rational interpretation of the observed action [40,41]. Here, we show that, rather than being restricted to external, environmental aspects of reality (e.g., a ball jumps an obstacle to reach a new location versus a ball jumps to reach a new location but there is no obstacle present), the situational constraints through which the rational attributes of an observed behavior are estimated, are extended to self-centred, sensorimotor properties that observers share with the observed agents.

This issue is currently debated in the literature. On one hand, previous findings suggest that in early infancy such sensorimotor cues do not play an essential, selective role in the rational interpretation of observed actions. For example, Southgate and colleagues [42] showed that 6- to 8- month-old infants attributed rational properties to observed actions even when the movements used to achieve them were biomechanically impossible. In their study, rationality was defined as conditions in which the observed goal-directed movements were adapted to external situational constraints, independently of the biomechanical plausibility of these movements. On the other hand, other evidence suggests that a rational interpretation of goal-directed actions may be predicated upon sensorimotor information conveyed by movement kinematics [43]. On a similar line, Southgate and co-workers [44,45] recently showed that the motor system of 9- to 15-months old infants was activated during the prediction of observed actions. The authors proposed that the activation of the motor system, instead of being driven by current visual information, was driven by the infants' expectations about the movements by which an attributed goal would likely be achieved. Given these contradicting data, one may speculate that the coupling of a rational interpretation of goal-directed actions with the processing of sensorimotor cues such as object affordances might be highly dependent on motor expertise acquired from experience [46]. Furthermore, this coupling might mature later in development. Our results suggest that the coupling of biomechanical with probabilistic priors may be particularly strong in adult observers, presumably equipped with a high degree of motor expertise.

Biomechanical and probabilistic priors may recruit two different – and parallel – neural systems that occasionally combine to derive information about tool use and function from observation. However, the exact nature and function of these systems is still a matter of conjecture. Effector-dependent, biomechanical priors may exert their influence on action prediction by differently weighting action alternatives within the motor repertoire of posterior frontal cortices such that certain actions become favored over others according to the biomechanical constraints of the motor effectors. This process of weighting action alternatives could be mediated by reciprocal inhibitory connections within the motor cortices, either by suppressing or increasing the activity of current competitors [47]. Occasionally, probabilistic priors may exert top-down influences on the selection of action alternatives within premotor cortices by using evidence gathered from past events to re-assigning new weights to the set of possible actions. Interestingly, these probabilistic priors may recruit more anterior frontal regions, such as the dorsolateral [48] or the inferior parts [49] of the dorsolateral prefrontal cortex. As a result, one may speculate that an abnormal connectivity between dorsolateral prefrontal and premotor regions – resulting from an impaired biasing influence from anterior to more posterior frontal cortices – would lead to abnormal action selection [50,51]. Such abnormal selection might jeopardize acquisition of motor expertise and the ability to infer other people's intentions from observation [52].

Conclusion

To our knowledge, the present study provides the first evidence that object affordances play a major role in the learning and prediction of observed tool-use behaviors. In particular, we show that perceiving observed behaviors as rational depends on low-level local cues from which their biomechanical costs are estimated with regard to their final goals. We suggest that biomechanical expectations elicited by affordances impede or bias the extraction of probabilistic regularities from past events. When these statistical regularities favor the observation of biomechanically suboptimal behaviors, biomechanical expectations delay the acquisition of probabilistic priors. Consequently, they also hinder the use of these priors in solving the uncertainty that is associated with incomplete visual signals.

Interestingly, one may extrapolate from our results that increasing the number of observations for suboptimal behaviors would further boost the weight devoted to probabilistic information in the participants' decisions. If this is the case what might this boost reflect and how might the brain represent it? Further studies should investigate how, and whether, the increasing weight of probabilistic information is associated with an update of biomechanical priors. Such an update could occur through a mechanism of visuomotor learning mediated by the plastic properties of the motor system [53-55]. This would allow one to determine whether the interaction between a 'rational' interpretation of actions and the detection of affordances recruits a modular, domain-specific process that would configure the experience of the external world *per se*. Implications for the social learning of tool use could be particularly important, as it would suggest that the larger the magnitude of this interaction for learners, the less able they would be to predict and learn from biomechanically suboptimal or unexpected behaviors. More generally, we believe that this cognitive selectivity for biomechanical optimality could contribute to the convergence of individual behaviors towards homogeneous patterns [17]. This could arise in the absence of high-level, faithful social transmission mechanisms such as true imitation of observed action goals and means [56- 58]. Affordances could enhance the efficiency of less precise, though less costly, forms of social learning strategies in the acquisition of novel tool use, like emulation learning [59] or stimulus enhancement [60].

References

1. St Amant R, Horton TE (2008) Revisiting the definition of animal tool use. Anim Behav 75: 1199-1208.
2. Seed A, Byrne R (2010) Animal tool-use. Curr Biol 20: 1032-1039.
3. Hernik M, Csibra G (2009) Functional understanding facilitates learning about tools in human children. Curr Opin Neurobiol 19: 34-38.
4. Johnson-Frey SH (2003) What’s so special about human tool use? Neuron 39: 201-204.
5. Osiurak F, Jarry C, Le Gall D (2010) Grasping the affordances, understanding the reasoning: toward a dialectical theory of human tool use. Psychol Rev 117: 517-540.
6. Vaesen K (2012) The cognitive bases of human tool use. Behav Brain Sci: In press.
7. Gergely G (2007) Learning “about” versus learning “from” other minds: natural pedagogy and its implications. In: Carruthers P, Laurence S, Stich S, editors. The innate mind: Vol3, Foundations and the future. Oxford, UK: Oxford University Press: pp 170-198.
8. Gergely G, Csibra G (2003) Teleological reasoning in infancy: the naive theory of rational action. Trends Cog Sci 7: 287-292.
9. Baker CL, Saxe R, Tenenbaum JB (2009) Action understanding as inverse planning. Cognition 113: 329-349.
10. Buchsbaum D, Gopnik A, Griffiths TL, Shafto P (2011) Children’s imitation of causal action sequences is influenced by statistical and pedagogical evidence. Cognition 120: 331-340.
11. Chambon V, Domenech P, Pacherie E, Koechlin E, Baraduc P et al. (2011) What are they up to? The role of sensory evidence and prior knowledge in action understanding. PloS One 6: e17133
12. Griffiths TL, Kemp C, Tenenbaum JB (2008) Bayesian models of cognition. In: Sun R, ed. The Cambridge handbook of computational psychology. Cambridge, UK: Cambridge University Press. pp 59-100.
13. Griffiths TL, Chater N, Kemp C, Perfors A, Tenenbaum JB (2010) Probabilistic models of cognition: exploring representations and inductive biases. Trends Cog Sci 14: 357-364.
14. Casler K, Kelemen D (2005) Young children's rapid learning about artifacts. Developmental Sci 8: 472-480.
15. Casler K Kelemen D (2007) Reasoning about artifacts at 24 months: the developing teleo-functional stance. Cognition 103: 120-130.
16. Csibra G, Gergely G (2007) « Obsessed with goals »: functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol 124: 60-78.
17. Jacquet PO, Tessari A, Binkofski F, Borghi AM (2012) Can object affordances impact on human social learning of tool use? Behav Brain Sci: In press.
18. Gibson JJ (1979) The ecological approach to visual perception. Boston, MA: Houghton Mifflin.
19. Norman DA (1988) The Psychology of Everyday Things. New York, NY: Basic Books.
20. Ellis R, Tucker M (2000) Micro-affordance: the potentiation of components of action by seen objects. Brit J Psychol 91: 451-471.

21. Borghi AM, Bonfiglioli C, Lugli L, Ricciardelli P, Rubichi S et al. (2007) Are visual stimuli sufficient to evoke motor information? Neurosci Lett 411: 17-21.
22. Grèzes J, Tucker M, Armony J, Ellis R, Passingham RE (2003) Objects automatically potentiate action: an fMRI study of implicit processing. Eur J Neurosci 17: 2735-2740.
23. Rosenbaum DA, van Heugten CM, Caldwell GE (1996) From cognition to biomechanics and back: the end-state comfort effect and the middle-is-faster effect. Acta Psychol, 94: 59-85.
24. Rosenbaum DA, Vaughan J, Barnes HJ, Jorgensen MJ (1992) Time course of movement planning: selection of handgrips for object manipulation. J Exp Psychol Learn 18: 1058-1073.
25. Sartori L, Straulino E, Castiello U (2011) How objects are grasped: the interplay between affordances and end-goals. PLoS One 6: e25203
26. Weiss DJ, Wark JD, Rosenbaum DA (2007) Monkey see, monkey plan, monkey do: the end-state comfort effect in cotton-top tamarins (Saguinus oedipus). Psychol Sci 18: 1063-1068.
27. Dennett DC (1982) Beyond beliefs. In: Woodfield A, ed. Thought and object: essays on intentionality. Oxford, UK: Oxford University Press. pp 1-97.
28. Dennett DC (1995) Darwin’s dangerous idea. New York, NY: Simon and Schuster.
29. Gregory RL (1981) Mind in science: a history of explanations in psychology and physics. Cambridge, UK: Cambridge University Press.
30. Clark A (1997) Being there: putting brain, body, and world together again. Cambridge, MA: MIT Press.
31. Sterelny K (2003a) Thought in a hostile world. New York, NY: Blackwell.
32. Sterelny K (2003b) Cognitive load and human decision, or, three ways of rolling the rock uphill. In: Carruthers P, Laurence S, Stich S, editors. The innate mind: Vol2, Culture and cognition. Oxford, UK: Oxford University Press. pp 218-233
33. Tessari A, Bosanac D, Rumiati RI (2006) Effect of learning on imitation of new actions: implications for a memory model. Exp Brain Res 173: 507-5013.
34. Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event-coding. Behav Brain Sci 24: 849-937.
35. World Medical Association Declaration of Helsinki (October 2008). http://www.wma.net/e/policy/b3.htm (accessed 16 August 2010)
36. Desmurget M, Prablanc C, Rossetti Y, Arzi M, Paulignan Y et al. (1995) Postural and synergic control of three-dimensional movements of reaching and grasping. J Neurophysiol 74: 905-910.
37. Paulignan Y, Frak VG, Toni I, Jeannerod M (1997) Influence of object position and size on human prehension movements. Exp Brain Res 114: 226-234.
38. Gergely G, Nádasdy Z, Csibra G, Bíró S (1995) Taking the intentional stance at 12 months of age. Cognition 56: 165-193.
39. Gergely G, Bekkering H, Király I (2002) Rational imitation in preverbal infants. Nature 415: 755.
40. Csibra G, Gergely G, Bíró S, Koós O, Brockbank M (1999) Goal attribution without agency cues: the perception of « pure reason » in infancy. Cognition 72: 237-267.

41. Onishi KH, Baillargeon R, Leslie AM (2007) 15-month-old infants detect violations in pretend scenarios. Acta Psychol 124: 106-128.
42. Southgate V, Johnson MH, Csibra G (2008) Infants attribute goals even to biomechanically impossible actions. Cognition 107: 1059-1069.
43. Kilner JM, Vargas C, Duval S, Blakemore SJ, Sirigu A (2004) Motor activation prior to observation of a predicted movement. Nat Neurosci 7: 1299-1301.
44. Southgate V, Johnson MH, El Karoui I, Csibra G (2010) Motor system activation reveals infants' on-line prediction of others' goals. Psychol Sci 21: 355-359.
45. Southgate V, Johnson MH, Osborne T, Csibra G (2009) Predictive motor activation during action observation in human infants. Biol Lett 5: 769-772.
46. Aglioti SM, Cesari P, Romani M, Urgesi C (2008) Action anticipation and motor resonance in elite basketball players. Nat Neurosci 11: 1109-1116.
47. Cisek P (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. Philos T Roy Soc B 362: 1585-1599.
48. Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302: 1181-1185.
49. Kilner J (2011) More than one pathway to action understanding. Trends Cog Sci 15: 352-357.
50. Barbalat G, Chambon V, Franck N, Koechlin E, Farrer C (2009) Organization of cognitive control within the lateral prefrontal cortex in schizophrenia. Arch Gen Psychiat 66: 377-86.
51. Barbalat G, Chambon V, Ody C, Domenech P, Koechlin E et al. (2011) Top-down control within the lateral prefrontal cortex in schizophrenia. Biol Psychiat 70: 73-80.
52. Chambon V, Barbalat G, Pacherie I, Jacquet PO, Franck N et al. (2011) Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia. Brain 134: 3725-38.
53. Catmur C, Walsh V, Heyes C (2007) Sensorimotor learning configures the human mirror system. Curr Biol 17: 1527-1531.
54. Stefan K, Cohen LG, Duque J, Mazzocchio R, Celnik P et al (2005) Formation of a motor memory by action observation. J Neurosci 25: 9339-9346.
55. Stefan K, Classen J, Celnik P, Cohen LG (2008) Concurrent action observation modulates practice-induced motor memory formation. Eur J Neurosci 27: 730-738.
56. Claidière N, Sperber D (2010) Imitation explains the propagation, not the stability of animal culture. P Roy Soc B-Biol Sci 277: 651-659.
57. Tennie C, Hedwig D, Call J, Tomasello M (2008) An experimental study of nettle feeding in captive gorillas. Am J Primatol 70: 584-93.
58. Acerbi A, Jacquet PO, Tennie C (2012) Behavioral constraints and the evolution of faithful social learning. Current Zoology 58: 307-318.
59. Acerbi A, Tennie C, Nunn C (2011) Modeling imitation and emulation in constrained search space. Learn Behav 39: 104-114.
60. Franz M, Matthews LJ (2010) Social enhancement can create adaptive, arbitrary and maladaptive cultural traditions. P Roy Soc B-Biol Sci 277: 3363-3372.

EXPERIMENTAL CONTRIBUTION

CHAPTER 3

The **Experimental Contribution Chapter 2** showed that prior information aquired from probabilistic exposure and prior information derived from an estimation of the biomechanical costs engaged during observed actions interact, and that this interaction modulates the prediction of participants regarding a demonstrator's intentions. In the study presented in the **Experimental Contribution Chapter 3**, I was interested in investigating whether such an interaction also translates into the motor system, possibly through a modulation of the motor resonance mechanisms.

To test this, I used single-pulse TMS over the primary motor cortex (M1) to probe corticospinal excitability changes (CSE) during an action prediction task that was similar to the one deigned in the **Experimental Contribution Chapter 2**. Thus, in this task, both the biomechanical optimality of actions being observed and their probability of occurrence were varied. The very similar stimuli were used: the task consisted of the presentation of a series of movies showing an actor performing two types of object-directed actions (lifting an object to open a box versus turning an object to switch-on a light) using either a power (whole hand grip) and a precision grip (pinch grip). In a first step, participants were presented a series of incomplete action movies where only the grasping phase was visible. In a second, step, they were presented a series complete action movies where both the grasping phase and the action's final outcome were visiable, and where one action type was more likely observed than another. In third step, a second series of incomplete action movies was presented. For each movie of the three series, participants were required to predict the final intention of the actor (open the box *versus* switch-on the light). Three groups of participants were tested, each being subjected to a specific probability distribution of optimal and suboptimal actions during the complete movie series. Single-pulse was assigned over the left M1 during the observation of incomplete action movies, prior and after the observation of complete action movies in which the probabilistic biases were implemented. Comparing MEPs between the first and the second series of incomplete movies enabled us to study the effect of probabilistic sampling of past events on CSE as well as its interaction with the visual information conveyed by the movement kinematics. By doing so, we were able to draw inference about whether the motor simulation processes were tuned by higher-order expectations, namely, the expectations generated by the probabilistic exposure to a specific action type.

It has been argued that single-pulse TMS offers a dynamic picture of motor simulation processes that are at stake during action observation (Fadiga et al., 2005). The reason is that the TMS application over M1 enables the on-line measurement of CSE at high temporal resolution. The first evidence of a modulation of CSE during the observation of biological actions in humans was provided by Fadiga and collaborators in 1995. In this seminal study, experimenters applied TMS over the cortical representation of the hand in M1 and recorded motor-evoked potentials (MEPs) of various hand muscles, while participants observed transitive (e.g., grasping an object) or intransitive (e.g., lifting the arm) arm-hand movements. During observation of grasping action, the amplitude of MEPs recorded from two muscles involved in grasping movements (the first dorsal interosseus ‘FDI’ and the opponent pollicis ‘OP’) increased, as compared with those observed in the control condition. According to the authors, this finding demonstrates the involvement of a motor simulation mechanism during action observation, a prerequisite for action understanding and prediction.

The study by Fadiga and collaborators provided the basis of more than 15 years of single-pulse TMS studies on action observation (Alaerts et al., 2009; Alaerts et al., 2010; Aziz-Zadeh et al., 2002; Brighina et al., 2000; Catmur et al., 2007; Cesari et al., 2011; Clark et al., 2004; Gangitano et al., 2001; Lago and Fernandez-del-Olmo, 2011). The underlying hypothesis of these studies is that observing others' actions potentiates the neural populations located in the area corresponding to the cortical representation of muscles that are involved in the execution of the observed actions. According to Fadiga and collaborators (2005), this potentiation would decrease the activation threshold of these neurons, which leads to a facilitation effect after the TMS delivery, as reflected by an increase of CSE. This facilitation effect is often interpreted as an indice that motor simulation mechanisms operate during action observation.

In parallel, single-pulse TMS over M1 has also been recently used to uncover, with action preparation paradigms, the impact of statistical regularities of past events over the CSE of participants (Bestmann, 2008; van Elswijk et al., 2007). It appears that the predictability of an event is encoded within the motor system, as indirectly revealed by an increase of CSE when cues serving for action preparation were highly expected by participants.

Thus, single-pulse TMS applied over M1 offers a powerful technique allowing the measurement of CSE changes that could occur through the modulation of sensorimotor components of action stimuli as well as the modulation of their probability distributions.

Modulating human motor resonance:

Exposure to suboptimal actions suppresses corticospinal excitability

(Preliminary draft)

Pierre O. Jacquet[1,2,3*], Alice C. Roy[2,4], Valérian Chambon[5], Anna M. Borghi[3,6], Alessandro Farnè[1,2], Karen T. Reilly[1,2]

[1] INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Center, ImpAct Team, Lyon, France;

[2] University Lyon 1, 69000 Lyon, France;

[3] Department of Psychology, Bologna University, Bologna, Italy;

[4] CNRS FRE 3406, Institut des Sciences Cognitives, L2C2, Bron, France.

[5] Laboratoire de Neurosciences Cognitives, INSERM, Paris, France

[6] Institute of Sciences and Technologies of Cognition, National Research Council, Rome, Italy.

Abstract

Motor resonance mechanisms – the automatic activation in the motor system of sensorimotor representations that are equally recruited during the execution and the observation of an action – are assumed to enable an observer to infer the goal or intention of another agent’s action. However, it is not yet clear whether, and how, these mechanisms can be adaptively modulated by different sources of prior information, such as i) the prior knowledge an observer may accumulate about another agent’s behavior through the probabilistic sampling of past events, ii) the biomechanical optimality of the observed actions (the muscular and articulator costs of given action). The present study aimed to test whether the interactions between biomechanical expectations and prior knowledge about a demonstrator’s behaviors, which are known to influence the participants’ predictions of the demonstrator’s intentions, can modulate the motor system activity. To test this, we used single-pulse TMS applied over the primary motor cortex (M1) to derive a well-established measure the corticospinal excitability (CSE) of participants during an action prediction task. In this task participants were required to infer, under various conditions of visual uncertainty, the intentions of a demonstrator performing tool-use behaviors. Both the probability of observing the demonstrator achieving biomechanically optimal and suboptimal actions was varied. Our results show that motor resonance processes are sensitive to biomechanical optimality, in such a way that they adaptively adjust their activity depending on prior expectations of the observer. This adjustment is here demonstrated via the maintenance of corticospinal excitability in conditions where biomechanical and prior knowledge acquired from probabilistic exposure strictly converge, and a decrease of excitability when they diverge. This regulatory activity could reflect an adaptive mechanism whereby the brain efficiently weights information gathered from probabilistic sampling of past observations to optimize the understanding, the prediction, and possibly the acquisition of new behaviors.

Keywords: action prediction, prior information, motor resonance, transcranial magnetic stimulation.

Introduction

The human cultural landscape abounds with unstable, changing, and open-ended behavioral environments (Boyd and Richerson, 1985). Under such circumstances, understanding, predicting and learning new behaviors from observation requires the ability to adaptively deal with each environment's intrinsic uncertainty by exploiting, adjusting, or inhibiting information of various sources (Collins and Koechlin, 2012). One type of information is the observer's expectations regarding the type of motor sequence performed to achieve a predicted goal given the *biomechanical* constraints of the effector itself (Gallese and Goldman, 1998). It has previously been shown that actions of both human and non-human animals aim to be biomechanically optimal, i.e. to minimize the muscular and articulator costs of the action (Flash and Hochner, 2005; Rosenbaum et al., 1992, 1996; Sartori et al., 2011; Weiss et al., 2007). Recent experimental work in humans suggests that this function of minimizing costs also transfers to action prediction (Jacquet et al., 2012b, in press), with observers expecting others to behave in accordance with biomechanical optimality rules. Biomechanical prior expectations play a role in action prediction when the observed action is directed to, or mediated by, an object or a tool, as the tool's affordances 'tell' the agent and the observer how to act on it appropriately. Affordances thus reduce the number of possible motor acts one can perform on an object or a tool given the biomechanical constraints of the agents (Gibson, 1979). Predicting and learning new behaviors from observation also depends on acquiring prior knowledge about another person's intentions, or goals, from *probabilistic* sampling of past experience (Baker et al., 2009; Chambon et al.,2011; Csibra and Gergely, 2007). Both biomechanical and probabilistic information have been demonstrated to be crucial for action prediction (see Chambon et al., 2011). Specifically, it has been shown that biomechanical and probabilistic priors adaptively interact during the prediction of optimal and suboptimal actions, with the nature of their interaction depending on whether subjects were exposed to probabilistic behaviors that converge with, or diverge from, biomechanical optimality (Jacquet et al., 2012b, in press). Indeed, the influence of biomechanical expectations on action prediction was found to progressively decrease over time, as the probability of biomechanically *suboptimal* behaviours concomitantly increased (Jacquet et al., 2012b, in press).

How the brain mediates competition between these two potentially conflicting sources of information remains unclear. Biomechanical and probabilistic priors may recruit two different – and parallel – neural networks that converge at some point in order to derive information from observation about action movements and goals. It has been suggested that probabilistic priors may exert top-down influences on the selection of action alternatives within premotor cortices, by using evidence gathered from probabilistic sampling of the environment (i.e., past observations) to re-assign new weights to the whole set of possible actions. Interestingly, such top-down influences would recruit anterior regions of the frontal hierarchy, such as the dorsolateral (Koechlin et al., 2003) or the inferior parts (Kilner et al., 2011) of the prefrontal cortex. In contrast, biomechanical priors may exert their influence on action prediction downstream of the frontal hierarchy by differentially weighting action alternatives within the motor repertoire of posterior frontal cortices, so that certain actions become favoured over others according to the biomechanical constraints of the motor effectors. This process of weighting action alternatives would be mediated by reciprocal inhibitory connections within the motor cortices, either by suppressing or increasing the activity of current competitors (Cisek et al., 2007). The emergence of biomechanical priors may thus be closely related to motor *resonance* mechanisms – i.e., the transformation of visual input containing kinematics of a biological movement into the corresponding motor programs stored in the observer's motor repertoire. This transformation is assumed to be mediated by mirror system activity of the human brain (Rizzolatti and Craighero, 2004). This mirror activity would allow the observer to understand the outcome of an action by matching the motor components of the observed action with the corresponding sensorimotor representations stored in her/his own behavioral repertoire (Wilson, Knoblich, 2005; Urgesi et al., 2010; Avenanti and Urgesi, 2010). Thus, motor resonance might be a simulation device that aids goal understanding by replicating the observed action in the observer's own motor system. Accordingly, the represented action should be 'executable', that is, consistent with the constraints of the observer's motor system (Csibra, 2007).

It has long been argued that motor resonance processes are automatically activated by the mere observation of biological movements. If this is the case, they would actively participate in action *perception*, a view that has received considerable empirical support in the last two decades (see for a recent review Rizzolatti and Sinigaglia, 2010). However, the role of these resonance processes in action *prediction* is still a subject of controversy. Indeed, predicting others' behaviors through motor resonance would be possible under conditions in which the observer owns a representation of the intended goal. That is to say, prediction would be possible if, and only if, the goal underlying the observed motor act is known in advance, or is familiar to the observer (Csibra, 2007; Kilner, 2011). Thus, motor resonance mechanisms would be particularly well suited for predicting goal-directed behaviors achieved in stable, familiar contexts (Aglioti et al., 2008), while being less suited to unfamiliar and open-ended ones (Csibra, 2007; Csibra and Gergely, 2007).

As such, enriching or degrading the familiarity of an action – by increasing the probability of observing either familiar and biomechanically optimal actions, or unfamiliar, biomechanically suboptimal actions – should modulate motor resonance processes, respectively, increasing or lowering corticospinal excitability, according to biomechanical optimality (Southgate et al., 2008, 2009). The present study aimed to test this assumption by investigating whether, and how, priors derived from probabilistic exposure influence motor resonance mechanisms in a task that required predicting ongoing, open-ended behaviors. To do so we used an on-line transcranial magnetic stimulation (TMS) technique during an action prediction task that has been described previously (Jacquet et al., 2012b, in press). Applying single-pulse TMS over the primary motor cortex (M1) is a well-established way to probe motor resonance mechanisms that are active during action observation (Alaerts et al., 2009; Alaerts et al., 2010; Aziz-Zadeh et al., 2002; Brighina et al., 2000; Catmur et al., 2007; Cesari et al., 2011; Clark et al., 2004; Gangitano et al., 2001; Lago and Fernandez-del-Olmo, 2011). Observing an agent performing an action facilitates neural activity in an area corresponding to the cortical representation of muscles that are involved in the execution of that same action. This facilitation effect is measured as an increase in corticospinal excitability (CSE), and is expected to signal the involvement of motor resonance mechanisms during action observation (Fadiga et al., 1995, 2005).

During the action prediction task, adult participants were required to infer the intentions of a filmed demonstrator acting on a tool that elicited two distinct affordances, each of which was biomechanically optimal for performance of a specific goal. The biomechanical optimality of tool-use actions as well as the probability of observing the demonstrator using an optimal versus suboptimal strategy were varied. Single-pulse TMS was applied over the left primary motor cortex (M1) of participants while they viewed videos in which they had to predict the demonstrator's final goal.TMS was applied before and after they watched a block of videos in which the probability of observing optimal or sub-optimal actions was manipulated (see Material and Methods, below). Three groups of participants were each assigned a specific probabilistic environment (bias): i) the demonstrator *equally* used optimal and suboptimal behaviors to achieve his intention; ii) the demonstrator preferentially used *optimal* behaviors, and iii) the demonstrator preferentially used *suboptimal* behaviors. This procedure allowed us to assess the effect of varying the contribution of the two classes of priors (probabilistic and biomechanical) to action prediction, together with investigating how such variations translate into changes of corticospinal excitability.

Material and Methods

Participants

Fifty-four healthy volunteers (29 women) aged 19-36 (mean = 24, SD = 4.2) took part in an action prediction task similar to the procedure used in Jacquet et al. (2012b, in press). All were right-handed, naive to the purpose of the experiment, and reported normal or corrected-to-normal visual acuity. All participants gave written informed consent and received payment for their participation in the study. Information about the experimental hypothesis was provided only after the experiment had been completed. The experimental protocol was performed with approval of the local Ethical Committee (Comité de Protection des Personnes SUD-EST IV, no. 2010-A01180-39) and in accordance with the Declaration of Helsinki (2008). All participants did not present any neurological, psychiatric, or other medical problems that are contraindicated for TMS (Wassermann, 1998).

General Procedure

Participants sat in front of a monitor on which 2000 ms action movies were displayed. The experimental session was divided into three blocks. In all three blocks participants watched videos featuring a demonstrator acting on a tool and were required to guess the goal of the demonstrator's action (for a similar procedure, see Jacquet et al., 2012b, in press). The tool consisted of a movable handle screwed onto the lid of a box. The handle offered two distinct affordances enabling the demonstrator to grasp the object with a power or a precision grip (see fig.1). Using either grip, the demonstrator could achieve two actions: opening the box by lifting the handle (intention A); switching on the light by rotating the handle (intention B) (see fig.1). In the first and third parts participants watched *incomplete* movies in which the action stopped 800ms after movement onset (at this time only the grip was apparent not the demonstrator's final goal) and the last displayed frame was presented on the screen for 1200ms. During these incomplete videos a single TMS pulse was applied after movie onset. In the second part, participants watched *complete* movies in which the action lasted until achievement of the goal (opening or turning), and both the grasp and the demonstrator's goal were apparent. No TMS was applied during these complete videos, and each participant was randomly assigned to one of three experimental groups each characterized by a specific probability distribution of optimal and suboptimal complete action movies. These complete movies were used to bias participants in favour of either optimal or suboptimal behavioral strategies. Of note is that each video (incomplete and complete) was unique: it was shown only one time per block.

Using the power grip to achieve the intention of opening the box by lifting the handle was low-cost, as was using the precision grip to achieve the intention of switching the lights on by turning the handle. These two combinations were identified as *optimal* behaviors (low biomechanical cost). The precision grip increased the cost of achieving the intention of opening the box, whereas the power grip increased the cost of achieving the intention of switching on the lights. These two combinations were identified as *suboptimal* behaviors (high biomechanical cost) (see Figure 1).

Participants assigned to the 'No bias' group had an equal probability of observing the demonstrator achieving his intention by performing an optimal or a suboptimal behavior. Participants assigned to the 'Convergent bias' group were biased towards 'optimal' behaviors, to the detriment of 'suboptimal' behaviors. In 80% of the 'box opening' trials the demonstrator opened the box using a power grip, and in 80% of the 'light switching' trials he switched on the lights using a precision grip. Here, behaviors that were preferentially used by the demonstrator *converged* towards the participant's biomechanical priors. Finally, participants assigned to the 'Divergent bias' group were biased towards 'suboptimal' behaviors, to the detriment of 'optimal' behaviors. In 80% of the 'box opening' trials the demonstrator opened the box using a precision grip, and in 80% of the 'light switching' trials he switched on the lights using a power grip. Here, the behaviors that were preferentially used by the demonstrator *diverged* from the participant's biomechanical priors. Of note is that the distributions of optimal and suboptimal behaviors during the incomplete movie block were kept equiprobable: participants were presented the same number of optimal and suboptimal actions.

For each of the 288 action movies that composed an experimental session (96 incomplete in part 1, 96 complete in part 2, 96 incomplete in part 3) participants were required to predict the demonstrator's intention by producing a vocal response ('A' for opening the box; 'B' for switching on the lights). Vocal responses were recorded via a microphone. Participants were instructed to make their response as soon as they thought they had enough visual information to produce an accurate response. Note however that both complete and incomplete movies ran for the full 2000ms independently of the timing of the subject's response.

Prior to the experiment participants were familiarised with the task by watching six incomplete movies and an unbiased block of twelve complete movies.

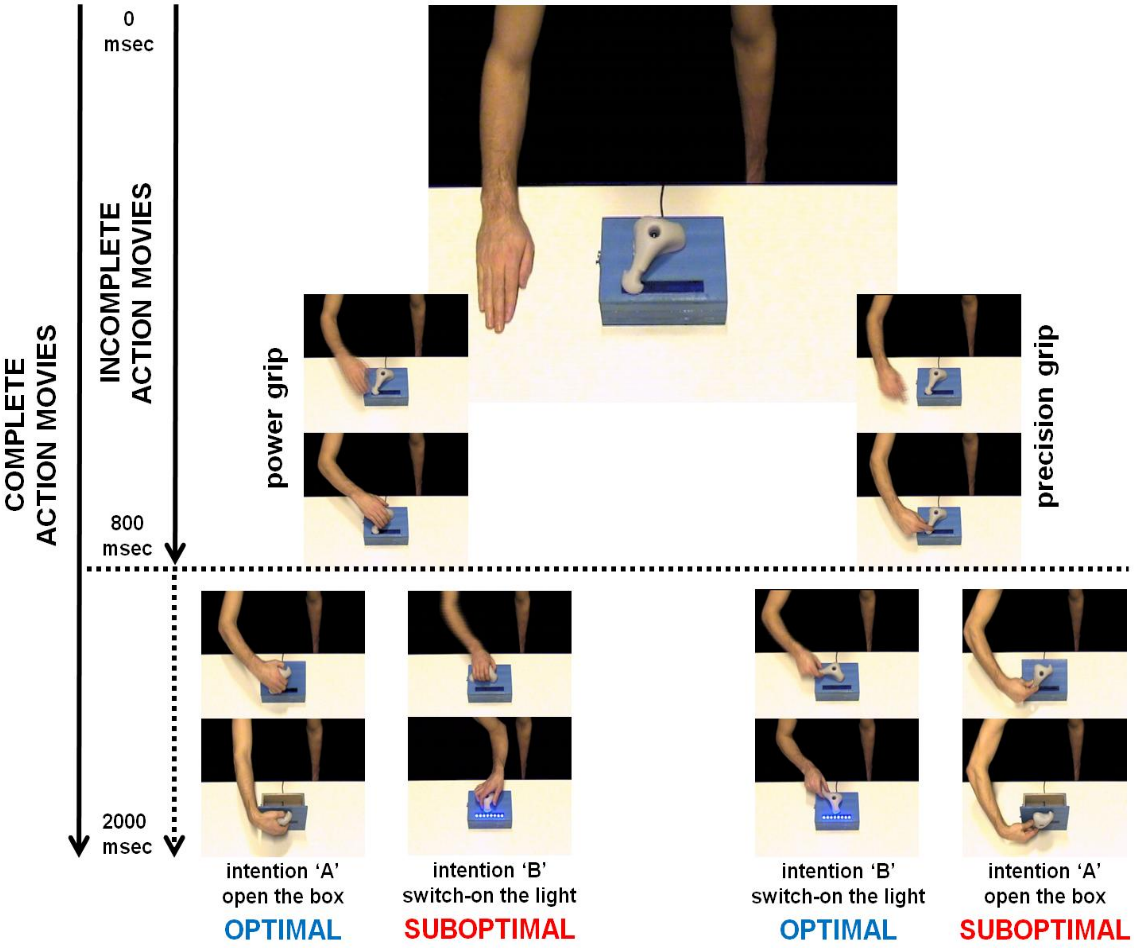


Figure 1: Examples of the four combinations 'grip × intention' that participants encountered during the experiment, and that lead to 'optimal' or 'suboptimal' behaviors. All combinations began with the demonstrator's static hand. The actor could then use either a 'power' or a 'precision' grip to achieve either the intention of opening the box or switching the lights on. The combination between the kind of grip and the kind of final intention resulted in the complete action as being labeled biomechanically optimal (OPTIMAL) or suboptimal (SUBOPTIMAL). Whereas the complete action movies lasted until the achievement of the underlying intention for a total duration of 2000ms, the incomplete action movies stopped 800ms after the movement onset (when the demonstrator was about to grasp the tool) while the last displayed frame remained on the screen for a duration of 1200 ms, so that observers had information about the grip but no information (on that trial) about the demonstrator's intention.

Typical trial

All trials started with a white fixation-cross that appeared for 2500ms on a dark background. The fixation cross was immediately followed by either a complete or an incomplete action movie that lasted 2000ms. On each trial participants verbally indicated which goal the demonstrator was about to achieve and at the end of the video response time was displayed on the screen for 500ms. For those trials in which participants did not respond, or responded too late, 'NO RESPONSE' was displayed on the screen. The next trial started immediately after the 500ms response feedback period. This feedback allowed us to avoid a 'guessing bias' that could occur during the presentation of complete action sequences, and that could hinder the integration of the probabilistic bias (see for a similar procedure, Jacquet et al., 2012b, in press).

For each incomplete movie, a single-pulse TMS was randomly delivered at 600, 700 and 800ms after the onset of the movies, when the type of grip used by the demonstrator was fully visible. Each block of incomplete movies included 16 trials without TMS in order to minimize the predictability of the stimulation. Each participant also performed three blocks of a TMS control condition, in which motor evoked potentials were recorded during 20 single TMS pulses over M1 while they viewed a white fixation cross located in the middle of a black screen (Gangitano et al., 2001). The inter-pulse interval was similar between 4800 and 5200 ms, and three blocks of 20 black-screen trials were included in the experiment; 1) before the first block of incomplete videos, 2) after the first block of incomplete videos, and 3) after the second block of incomplete videos. This procedure allowed us to check for any modulation of corticospinal excitability that did not depend on our experimental variables (e.g., fatigue).

The presentation of stimuli, the recording of vocal responses (response times) and the TMS triggering was synchronized using *Presentation* software (Neurobehavioral Systems, Inc, USA).

TMS and electromyographic (EMG) recording

Motor-evoked potentials (MEPs) were recorded from the first dorsal interosseous (FDI) muscle of the right hand. Of note is that the FDI muscle is strongly involved in the execution of each action type that was presented to participants. EMG recordings were performed using Ag-AgCl electrodes placed in a belly-tendon configuration. EMG activity was amplified and digitized with a CED Power 1401 interface and sampled at 5 kHz. Spike2 software (Cambridge Electronic Design, Cambridge, England) was used for off-line data analysis.

A Magstim rapid^2 stimulator (The Magstim Company, Carmarthenshire, Wales) generated single-pulse stimuli, delivered through a figure-of-eight coil (70 mm diameter) placed tangentially to the scalp with the handle pointing backward and at a 45° angle away from the midline. During the recording sessions, the coil was positioned over the left primary motor cortex (M1) in correspondence with the optimal scalp position (OSP), defined as the position from which MEPs with maximal amplitude were recorded from FDI. The OSP was identified by moving the intersection of the coil in 1cm steps around the hand area of the left motor cortex and by delivering TMS pulses at constant supra-threshold intensity. Participants wore a bathing cap on which the OSP for stimulation was marked. The mark and the coil position relative to the mark were then recorded on each participant's scalp using the SofTaxic

Navigator system (EMS, Italy). The coil was held by hand and its position with respect to the target on the reconstructed brain was continuously monitored during the experiment.

The individual resting motor threshold (rMT) of each participant was identified as the minimal stimulation intensity that produced motor evoked potentials (MEPs) of at least 50 μV in the FDI, with 50% probability on 10 consecutive stimulations (Rossini et al., 1994). Stimulation intensity during the recording sessions was set at 120% of the rMT and ranged from 33% to 50% (mean = 41%; SD = 6) of the maximum stimulator output. Using this procedure, a clear and stable signal was obtained from the targeted muscle in all participants.

Data preprocessing

Peak-to-peak amplitudes of the MEPs were determined off-line from the raw EMG data (Spike2, version 7.02). To control for background EMG activity the data were visually inspected during the experiment. MEPs the amplitude of which could not be distinguished from the background activity were removed from the analysis. Background EMG was further controlled by computing the root mean squared (RMS) of the EMG during the 100ms prior to the TMS pulse.. Trials on which the pre-stimulus RMS exceeded the average (of the 80 TMS trials in the block) by more than two standard deviations were discarded. Trials on which MEP amplitude was greater than or less than 2 standard deviations from the mean were also removed.

The peak-to-peak amplitudes (mV) of the remaining MEPs were then determined and converted into a corticospinal excitability (CSE) change index, expressed as the LOG(10) transformation of percentages of variation relative to the mean MEP amplitude obtained in the

TMS control conditions. Importantly, the CSE change indices collected during the first incomplete movie block were computed relative to the mean MEP amplitude of the control blocks 1 and 2; while the CSE change indices collected during the second incomplete movie block were computed relative to the mean MEP amplitude of the control blocks 2 and 3. This procedure allowed us to disentangle corticospinal excitability changes due to our independent variables from non-specific modulations in corticospinal variability across time.

Because we stimulated at 120% of resting motor threshold, the absolute amplitude of the control MEPs varied considerably between subjects. Since these control MEPs were used to normalize data collected during the incomplete movies we reasoned that it would be best to maximize the physiological homogeneity of all participants and to include only those participants who had mean MEP amplitudes during the control conditions greater than 0.8 mV. The choice of this criterion was based upon the now-common procedure of choosing a stimulation intensity that gives a 1mV MEP. Consequently, 11 participants were selected for the ‘No bias’ session analysis (7 women; mean age = 23, SD = 3.15), 10 for the ‘Convergent’ session analysis (3 women; mean age = 24.6, SD = 4.93), and 12 for the ‘Divergent’ session analysis (6 women; mean age = 24.3, SD = 3.92).

Behavioral performance. We analysed the percentage of correct responses (hits) and response times (RTs) collected for both complete and incomplete action movies. Participants did not receive any feedback about the accuracy of their responses during the incomplete action videos, and responses were classed as correct if the predicted goal was the same as that actually achieved by the demonstrator in the complete format of that particular video. All statistical analyses were performed separately for each of the three movie blocks.

The hit rates and RTs collected during the complete movie block were analysed using 2 × 2 × 3 repeated-measures ANOVAs. The first two-level factor was the 'type of behavior' (optimal versus suboptimal), the second two-level factor was the 'type of grip' (power versus precision), and the third, three-level factor was the 'probabilistic bias' (no bias versus convergent bias versus divergent bias).

The hit rates and RTs collected during the two incomplete movie blocks were then analysed using 2 × 2 × 2 × 3 repeated-measures ANOVAs. The first two-level factor was the 'block' (pre-test versus post-test), the second two-level factor was the 'type of behavior' (optimal versus suboptimal), the second two-level factor was the 'type of grip' (power versus precision), and the third, three-level between-subject factor was the 'probabilistic bias' (no bias versus convergent bias versus divergent bias).

MEP data. MEP data were only collected during incomplete videos, for which there was no information concerning the goal of the action. Since we were interested in knowing whether corticospinal excitability was modulated as a function of action prediction, a trial was classified as optimal or suboptimal according to the participant's decision on that trial, regardless of whether or not this decision was correct. Thus, 'optimal' and 'suboptimal' trials contain both hits and false alarm. The percentage of MEPs excluded (too much background EMG activity or MEPs that were 1.96 SD up or down the mean MEP amplitude) ranged between 7 and 9% for the pre- and post- blocks (84 trials each) and was approximately equal for the three different groups.

A 2 × 2 × 2 × 3 repeated-measures ANOVA was then performed, with 'block' (pre-test versus post-test), 'type of behavior' (optimal versus suboptimal) and 'type of grip' (power versus precision) as two-level within-subject factors and the ‘probabilistic bias’ (no bias versus convergent bias versus divergent bias) as three-level between-subjects factor.

Results

Behavioral performance (Hits and RTs)

Complete action movies

Figure 2 shows the mean percentage of correct responses (upper panel) and mean reaction times (bottom panel) collected during complete action movies. The 2 (type of behavior) × 2 (type of grip) × 3 (probabilistic bias) repeated-measures ANOVAs revealed a main effect of the ‘type of behavior’ on both hits ($F_{1,30} = 28.20$, $p < .001$) and RTs ($F_{1,30} = 43.79$, $p < .001$). Participants were more accurate and faster at predicting optimal than suboptimal behaviors (hits: 90% vs. 75%; RTs: 1374ms vs. 1449ms). A main effect of the ‘type of grip’ was found only for RTs ($F_{1,30} = 9.84$, $p < .01$), with participants being overall faster at predicting behaviors that were performed with a precision than a power grip (1398ms vs. 1425ms).

The two-way interaction (type of behavior) × (probabilistic bias) (see figure 2) was also significant for both hits ($F_{2.30} = 12.29$, $p < .001$) and RTs ($F_{2.30} = 26.74$, $p < .0001$). Post-hoc comparisons (LSD Fisher tests) indicated that during the no bias group – where both types of behaviors were equally probable – participants were more accurate (92% vs. 77%, $p < .05$) and faster (1398ms vs. 1425ms, $p < .05$) at predicting optimal compared to suboptimal behaviors. A similar though stronger pattern was observed in the convergent bias group where optimal behaviors were more likely to be observed. Participants were more accurate (96% vs. 59%, $p < .0001$) and faster (1307ms vs. 1510ms, $p < .05$) at predicting optimal compared to suboptimal behaviors. Furthermore, when compared to the no bias group, participants in the convergent bias group were less accurate and slower at predicting the unbiased suboptimal behaviors (hits = 59% vs. 77%; RTs = 1510ms vs. 1425ms). In the divergent bias group no differences were found between the optimal and suboptimal behaviors, despite the fact that the latter were more frequently shown than the former (hits = 84% vs. 86%, $p > .05$; RTs = 1428ms vs. 1415ms, $p > .05$). Thus, increasing the probability of observing suboptimal behaviors did not significantly increase the number of correct responses for these behaviors compared to the optimal ones, however it cancelled the natural preference towards optimal behaviors that was present in the no bias group.

Finally, the (type of grip) × (type of behavior) interaction was significant for RTs only ($F_{2.30} = 4.78$, $p < .05$). Overall, participants were faster at predicting suboptimal behaviors when achieved with a precision than with a power grip (1426ms vs. 1471ms, $p > .001$).

No other significant main or interaction effects were observed (all $p > .07$).

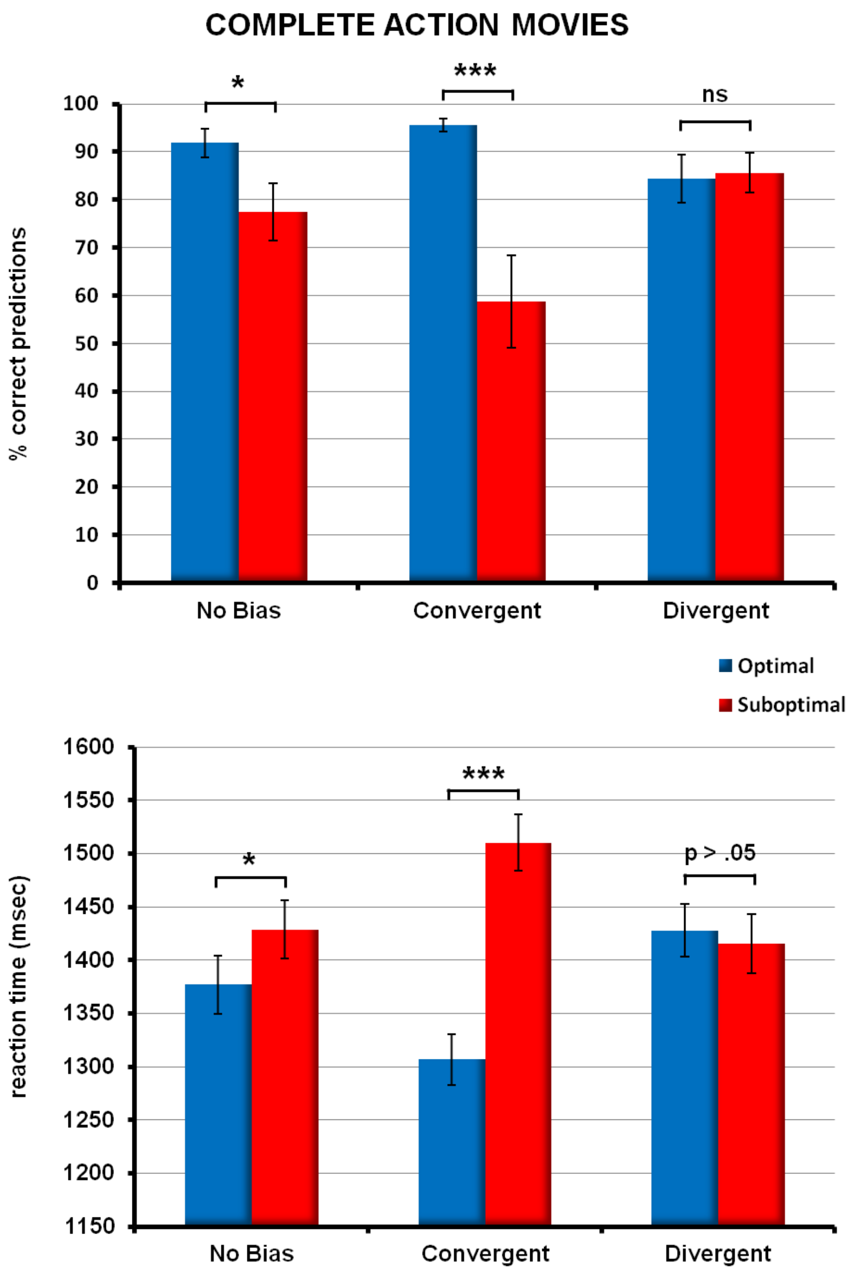


Figure 2. Behavioral performances. The upper graph represents the mean percentages of correct responses collected during complete action movies for all three sessions. The inferior graph represents the mean response times collected during complete action movies for all three sessions. The blue columns refer to the mean percentages of correct predictions for observed ‘optimal’ behaviors (pooled across ‘power’ and ‘precision’ grip). The red columns refer to the mean percentages of correct predictions for observed ‘suboptimal’ behaviors (pooled across ‘power’ and ‘precision’ grip). Error bars denote the standard error of the mean.

Incomplete action movies

Figure 3 shows the mean percentage of correct responses (upper panel) and mean reaction times (bottom panel) for the incomplete action movie blocks before (pre) and after (post) exposure to one of three probabilistic biases. The 2 (block) × 2 (type of behavior) × 2 (type of grip) × 3 (probabilistic bias) repeated-measures ANOVAs revealed a main effect of the 'type of behavior' on both hits ($F_{1.30} = 73.89$, $p < .001$) and RTs ($F_{1.30} = 37.72$, $p < .001$); participants were more accurate and faster at predicting optimal than suboptimal behaviors (hits: 64% vs. 37%; RTs: 1329ms vs. 1392ms). This preference for optimal behaviors significantly differed from chance (t-test compared to 50%, $t > 8.68$, $p < .001$). The main effect of 'type of grip' was significant for RTs only ($F_{1,27} = 15.31$, $p < .001$). Participants were faster at predicting behaviors performed with a precision grip than a power grip (1337ms vs. 1382ms, $p < .001$).

The two-way interaction 'type of behavior' × 'type of grip' was significant for hits only ($F_{1,30} = 6.24$, $p < .05$). Post-hoc tests (LSD Fisher tests) showed that participants were more accurate at predicting optimal rather than suboptimal behaviors, for both power (61% vs. 39%, $p < .001$) and precision grips (67% vs. 34%, $p < .001$).

Finally, the three-way interaction 'block' × 'type of behavior' × 'probabilistic bias' (see figure 3) was significant for hits ($F_{2,30} = 4.51$, $p < .05$). First of all, post-hoc comparisons (LSD Fisher tests) indicated that accuracy in pre-test was similar for the three probabilistic bias groups for both optimal and suboptimal behaviors (no bias = 64% vs. 35%; convergent bias = 68% vs. 39%; divergent bias = 62% vs. 38%; all $p > .35$). This is crucial as it demonstrates that the predictions of each group were equally guided *by default* by biomechanical expectations conveyed by the object's affordances. Furthermore, in the no bias group, no differences were observed between the pre-test and the post-test, for either optimal (64% vs.63%, $p > .05$) or suboptimal behaviors (35% vs. 35%, $p > .05$), and participants were more accurate at predicting optimal than suboptimal behaviors in both the pre (64% vs. 35%, $p < .001$) and post-test blocks (63% vs. 35%, $p < .001$). Of note is that, in the two blocks, this preference for optimal behaviors significantly differed from chance (pre-test = t-test compared to 50%, $t > 4.28$, $p < .01$; post-test = t-test against 50, $t > 4.54$, $p < .01$). In the convergent group, a similar pattern was observed for optimal behaviors, with participants being equally accurate in pre-test and in post-test (68% vs. 72%, $p > .05$). However, the accuracy decreased in post-test for suboptimal behaviors (39% vs. 29%, $p < .05$). Note that in both these blocks, participants were more accurate at predicting optimal than suboptimal behaviors (pre-test = 68% vs. 39%, $p < .001$; post-test = 72% vs. 29%, $p < .001$). Once again, in the two blocks of incomplete movies, the preference for optimal behaviors significantly differed from chance (pre-test = t-test for single mean, $t > 6.65$, $p < .001$; post-test = t-test compared to 50%, $t > 7.17$, $p < .001$). Post-hoc analyses also showed that the probabilistic exposure to optimal behaviors strengthened the participants' preference for these behaviors in the post-test, when compared to the no bias group in which both behaviors were equally presented (72% vs. 63%, $p < .05$). Interestingly, in the divergent group, the preference for optimal behaviors observed in pre-test (62% vs. 38%, $p < .001$) which differed significantly from chance level (t-test compared to 50%, $t > 3.80$, $p < .0$) was cancelled in the post-test, as indicated by the fact that performance for both optimal (56%; t-test against 50, $t > 1.89$, $p > .05$) and suboptimal behaviors (43%; t-test compared to 50%, $t > -1.52$, $p > .05$) did not differ from chance. A comparison with the post-test of the no bias session further shows that after being exposed to the divergent bias, participants were significantly more likely to predict suboptimal behaviors (35% vs. 43%, p < . 05). This result indicates that the probabilistic exposure to suboptimal behaviors conflicted with biomechanical expectations of participants in such a way that the preference for optimal behaviors was no longer observed.

No other significant main effect effects were observed (all p > .08).

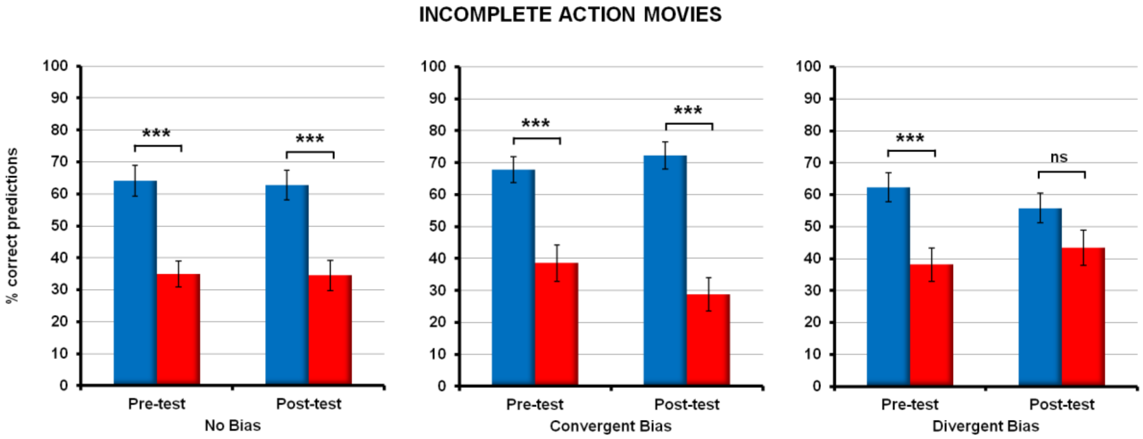


Figure 3. Behavioral performances. The upper graphs represent the mean percentages of correct responses collected during incomplete action movies of the pre-test and post-test, and for all three sessions. The blue columns refer to the mean percentages of correct predictions for observed 'optimal' behaviors (pooled across 'power' and 'precision' grip). The red columns refer to the mean percentages of correct predictions for observed 'suboptimal' behaviors (pooled across 'power' and 'precision' grip). Error bars denote the standard error of the mean.

Behavioral performances: preliminary discussion

These results demonstrate that, consistent with previous findings (Chambon et al., 2011; Jacquet et al., 2012b, in press), participants in both the convergent and divergent groups successfully integrated the probability distributions. This was evidenced by comparing the results for complete movies for the two biased groups with those of the no bias group (see figure 2). Results from the no bias group are characterized by a natural preference for optimal behaviors, that is, for behaviors that satisfied the participants' biomechanical prior expectations. Indeed, in the absence of any relevant information about which behaviour was most likely to be performed (in the no bias group the demonstrator achieved his intention by equally performing either optimal or suboptimal behaviors), participants made predictions on the basis of their biomechanical priors, even though there was enough visual information conveyed by the action scene (the video) to predict with 100% accuracy using only visual information. Note that this pattern was reinforced in the convergent bias group where *optimal* behaviors were more likely to be observed, as indicated by an increase in reaction times and a decrease in hit rate for the unbiased, suboptimal behaviors. Conversely, this intrinsic preference for optimal behaviors disappeared in the divergent bias group in which *suboptimal* behaviors were more likely to be observed.

These observations further highlight the impact of biomechanical expectations on the performances: they were overall more accurate and faster at predicting behaviors that minimized biomechanical costs, irrespective of probabilities. Indeed, in both the complete (see figure 2) and incomplete (see figure 3) action movies of the no bias group (i.e. who observed the demonstrator selecting between the two available behavioral strategies with equal probability), participants preferentially chose intentions achieved by optimal actions rather than suboptimal actions. This result demonstrates that when participants cannot rely on prior knowledge acquired from recent experience (i.e., on probability) to decide how an observed agent is most likely to behave, they rely on their biomechanical priors *by default*. That is, they assume that the observed agent behaves ‘optimally’, favouring strategies that minimize biomechanical costs.

Finally, the present results show that the magnitude of the probabilistic bias differed as a function of the type of behavior, with participants' biomechanical expectations overriding the effect of the probabilistic bias. In the convergent bias group (probabilistic bias assigned to optimal behaviors), performance was facilitated for the optimal behaviors, as expected. The fact that such a pattern was observed in both the complete movies (see figure 2) and the post-test blocks of incomplete movies (see figure 3) suggests that it is particularly costly for participants to take decisions against what is in line with rules of biomechanical optimality, even though the amount of visual information is sufficiently high to infer the demonstrator's intention based on movement kinematics alone. Strikingly, a very different pattern emerged for the *divergent* bias group. Indeed, in this group no difference was found between the optimal and suboptimal behaviors, although the latter were more frequently observed than the former. This pattern suggests that participants actively integrated both types of priors: when probabilistic and biomechanical priors diverged, the overall effect tended to sum to zero, resulting in performance that did not significantly differed from chance for both optimal and suboptimal behaviors.

MEP data (CSE change index)

Incomplete action movies

The 2 (block) × 2 (type of behavior) × 2 (type of grip) × 3 (probabilistic bias) repeated-measures ANOVAs revealed a main effect of the 'block' on the CSE change index ($F_{1.30}$ = 6.92, $p < .05$). Corticospinal excitability was greater in the pre-test compared to the post-test blocks (139% vs. 116%), independent of the type of probabilistic exposure.

The most interesting finding was the significant two-way interaction effect ‘block’ × ‘probabilistic bias’ ($F_{2,30}$ = 5.66, p = .01). As expected, post-hoc comparisons (LSD Fisher tests) showed that CSE change index for the pre-test block did not differ between the three probabilistic bias groups (no bias = 141%; convergent = 128%; divergent = 145%; all p > .45). Critically, in the divergent session, CSE was significantly reduced in the post-test block (145 vs. 100, p < .001). A similar, albeit not significant tendency was present in the no bias group, CSE being slightly lower in the post-test block (141% vs. 114%, p=.062) whereas, in the convergent group, MEP amplitude had the opposite, non-significant tendency (128% vs. 138%, p > .31). These results indicate that the type of probabilistic environment that participants were exposed to prior to the post-test block affected the CSE. In particular, only when this environment departed from that fitting biomechanical priors, the CSE level was significantly decreased.

No other significant main interaction effect were observed (all p > .29).

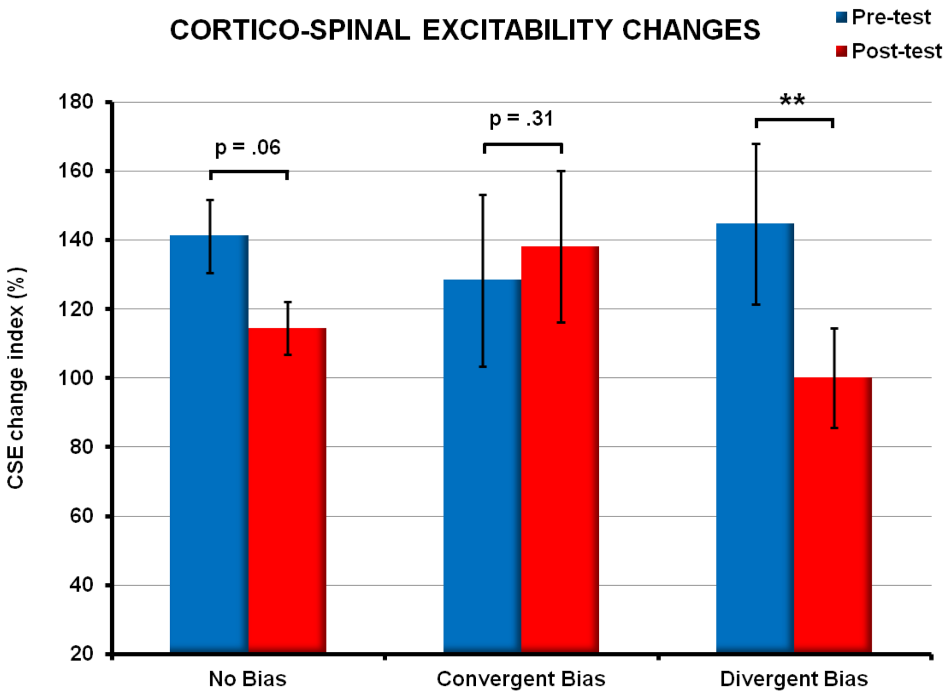


Figure 4. MEPs data. The graph represents the mean CSE change index recording during the incomplete action movies for all three sessions. The CSE change index reflects the % of variation of MEPs amplitude relative to the mean MEPs amplitude recorded during the TMS control condition (see table 1). The blue columns refer to the mean CSE change index for the pre-test (pooled across 'power' and 'precision' grip as well as 'optimal' and 'suboptimal' behaviors). The red columns refer to the mean CSE change index for the post-test (pooled across 'power' and 'precision' grip as well as 'optimal' and 'suboptimal' behaviors). Error bars denote the standard error of the mean.

MEP data: preliminary discussion

Our results revealed a gradient of CSE that is a function of the adequacy of the probabilistic context sampled from the complete movies to the participants' biomechanical priors. When the probabilistic context satisfies biomechanical optimality rules, the CSE level is maintained, while when the probabilistic context contradicts or violates biomechanical optimality, the CSE level decreased. In the no bias group there was a slight (non significant) decrease in CSE, independently of the type of behaviors (see figure 4). Interestingly, in the convergent session there was a stabilization of the CSE relative to the post-test block. Finally, in the divergent session, where participants were most likely to observe behaviors that departed from biomechanical optimality, the level of CSE was significantly decreased post-test. Overall, these results show that CSE changes may depend on the type of behaviour that was most frequently shown to the subject (optimal versus sub-optimal) during the complete video session, and on the degree to which the probabilistic environment matched the participants' biomechanical expectations. Indeed, the divergent session profoundly suppressed CSE possibly because participants were maximally exposed to actions that did not satisfy biomechanical optimality rules.

General discussion

The objective of this study was to test whether the interactions between biomechanical expectations (conveyed by tool affordances) and prior knowledge (acquired from probabilistic exposure) about a demonstrator's behaviors – known to influence the participants' predictions of the demonstrator's intentions – can modulate the motor system activity.

In a previous study (Jacquet et al., 2012b, in press), we provided the first evidence that biomechanical priors emerging from the detection of object affordances interact with priors acquired from probabilistic sampling of past events and can bias the prediction of observed actions. The behavioral data of the present study closely replicate our previous findings. We show that biomechanical expectations arising from the detection of object affordances adjust the participants' dependence on prior knowledge induced by probabilistic exposure. More specifically, when the behaviors gathered from probabilistic exposure satisfied these biomechanical prior expectations, participants efficiently combined both biomechanical and probabilistic priors to make their predictions. Conversely, when the behaviors gathered from probabilistic exposure violated the biomechanical priors, participants had to deal with two conflicting sources of prior information, which resulted in predictions close to chance level.

Overall, the tendency to the decrease in CSE level in the no bias session was suggestive of a general attenuation of the motor cortical activity. However, although CSE level significantly changed after repeated exposure to action stimuli, both the direction and the magnitude of this change tended to vary according to the probabilistic environment to which the participants were exposed, as well as on the degree to which this environment was biomechanical optimal. Thus, instead of reflecting an effect of neural adaptation, the suppression of corticospinal excitability observed as a tendency in the no bias, and clearly in the divergent session, are indicative of a change that does not allow the motor system to resonate anymore. This process occurred when the observed behavior departed from biomechanical optimality. Thus, in the 'convergent bias' group, participants watched the demonstrator acting in accordance with such expectations, and no CSE change was observed. In the 'divergent bias' group, the behaviors performed by the demonstrator frequently departed from biomechanical optimality, and the CSE level that was subsequently measured dramatically decreased, when compared to the initial exposure to incomplete actions (before the probabilistic bias). This pattern of findings suggests that motor resonance is not fully encapsulated and automatic, but can be actually quite malleable and sensitive to changes in the probabilistically-induced expectations. Further, we found that the magnitude of this inhibitory process was a function of the degree of reliability of the observer's biomechanical expectations, i.e., of how reliable biomechanical optimality rules were to predict the behavioral environment.

This CSE suppression may further suggest that the involvement of motor resonance mechanisms in action prediction is dependent on the interaction between the biomechanical prior expectations and the nature of the environment in which the prediction needs to be made, such that high levels of CSE are maintained only in environments that are consistent with these expectations. This is evidenced by the observation of different levels of CSE within each of the three environments (no bias vs. convergent vs. divergent). Motor resonance mechanisms – that are sensitive to biomechanical optimality – are therefore not affected by observation indistinctly (see also Catmur et al, 2007; Stefan et al., 2005, 2008). Rather, they could be modulated according to whether observation did, or did not, satisfy rules of biomechanical optimality. These findings are consistent with recent studies demonstrating that activity of the motor system of 9- to 15-months old infants is driven, during action prediction, by prior expectations about which movement is most likely to be performed in order to achieve a given goal, rather than by the current visual information alone (Southgate et al., 2009, 2010).

What may be the function of such a regulatory activity of the motor resonance mechanism? We suggest it may protect the observer against errors of prediction in the case where observation and expectations conflict, and would do so by leaving inadequate action representations with reduced resonance weights. In the context of the present task, such a weight reduction would operate during the observation of complete movies by down-regulating (in the no bias and the divergent bias groups) action representations that match with biomechanical prior expectations. Interestingly, the dorsal fronto-median cortex has been shown to underlie such ‘refraining’ process in a task requiring voluntary inhibition of prepotent impulsive responses, through top-down inhibition of premotor areas (Khün et al., 2008). Our results open the interesting possibility that a similar veto on action prediction also operates at a low, implicit level, with the aim of preventing observers from making maladaptive decisions. Thus, in the divergent bias session, where biomechanical optimality and probabilistic likelihood of the observed behaviors conflicted, participants might have refrained from inferring optimal behaviors. This regulation of motor resonance mechanisms may arise from top-down influences generated by the accumulation of probabilistic priors in the frontal regions. It could be that increasing the probabilistic likelihood of conflicting behaviors should in turn attenuate the influence of biomechanical expectations, through possible top-down adjustment of activation thresholds in the motor mirror system. This adjustment would result in decreasing the influence of motor resonance mechanisms on prediction, hence down-regulating the predictive value of low-level, kinematic features of the observed action. The role of motor resonance in action understanding and prediction has long been documented and debated (Rizzolatti and Craighero, 2004; Rizzolatti and Sinigaglia, 2010; for review). Through internal simulation of an action performed by a third party, an observer would be able to predict – on the basis of his or her own motor representations – the future states of the observed action, and, by extension, to infer the underlying intention (Gallese and Goldman, 1998). However, in order to be a reliable source of information for action prediction, motor resonance mechanisms must primarily be fed with a prior representation of the goal to predict (Csibra, 2007; Kilner, 2011). On this account, processes of motor resonance would not be suitable for inferring new goals, and would adjust poorly to unfamiliar, open-ended, environments. Regulation of automatic resonance processes through higher-order probabilistic representations of the environment may provide an adaptive mechanism to enable acquisition of unexpected, new behaviors. Interestingly, behaviors that over-ride rules of biomechanical optimization are regularly promoted by human culture. This is the case for many human cultural praxes, such as in some forms of sports, dances or music. Such praxes are often considered as socially valuable behaviors precisely because they are particularly difficult to learn and master. In some kinds of sport and dance practices, for example, biomechanical ‘suboptimality’ is perceived as a marker of excellence, and is socially rewarded for that reason. Importantly, vetoing direct, automatic resonance processes may also facilitate the learning of new behaviors from observing a non-expert model. Indeed, relying on prior knowledge gathered from probabilistic sampling of past observations may occasionally prove more helpful than merely evaluating the (biomechanical) optimality of the observed behavior.

Conclusion

Our results support the idea that motor resonance processes, sensitive to biomechanical optimality, may adaptively adjust their activity depending on prior expectations of the observer. This adjustment is here demonstrated via a maintenance of corticospinal excitability in conditions where biomechanical and probabilistic expectations strictly match, and a decrease of excitability when they diverge. This regulatory activity would reflect an adaptive mechanism whereby the brain efficiently weights information gathered from probabilistic sampling of past observations, to optimize the understanding, the prediction, and possibly the acquisition of new behaviors.

References

Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2010). Observing how others lift light or heavy objects: which visual cues mediate the encoding of muscular force in the primary motor cortex? *Neuropsychologia*, *48*(7), 2082-2090. doi:10.1016/j.neuropsychologia.2010.03.029

Avenanti, A., & Urgesi, C. (2011). Understanding « what » others do: mirror mechanisms play a crucial role in action perception. *Social cognitive and affective neuroscience*, *6*(3), 257-259. doi:10.1093/scan/nsr004

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current biology: CB*, *17*(17), 1527-1531. doi:10.1016/j.cub.2007.08.006

Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P., & Farrer, C. (2011). What are they up to? The role of sensory evidence and prior knowledge in action understanding. *PloS one*, *6*(2), e17133. doi:10.1371/journal.pone.0017133

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS biology*, *10*(3), e1001293. doi:10.1371/journal.pbio.1001293

Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press.

Csibra, G., & Gergely, G. (2007). « Obsessed with goals »: functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, *124*(1), 60-78. doi:10.1016/j.actpsy.2006.09.007

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13494-13499.

Fadiga, L, Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, *73*(6), 2608-2611.

Flash, T., & Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Current opinion in neurobiology*, *15*(6), 660-666. doi:10.1016/j.conb.2005.10.011

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gangitano, M., Mottaghy, F. M., & Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, *12*(7), 1489-1492.

Henson, R. N. A., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263-270.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357. doi:10.1016/j.tics.2011.06.005

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science (New York, N.Y.)*, *302*(5648), 1181-1185. doi:10.1126/science.1088545

Khün, S., Haggard, P., Brass, M. (2009). Intentional inhibition: how the "veto-area" exerts control. Human Brain Mapping, 30(9), 2834-28-43.

Lago, A., & Fernandez-del-Olmo, M. (2011). Movement observation specifies motor programs activated by the action observed objective. *Neuroscience letters*, *493*(3), 102-106. doi:10.1016/j.neulet.2011.02.014

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews. Neuroscience*, *11*(4), 264-274. doi:10.1038/nrn2805

Rosenbaum, D. A., Vaughan, J., Barnes, H. J., & Jorgensen, M. J. (1992). Time course of movement planning: selection of handgrips for object manipulation. *Journal of experimental psychology. Learning, memory, and cognition*, *18*(5), 1058-1073.

Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., Dimitrijević, M. R., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalography and clinical neurophysiology*, *91*(2), 79-92.

Southgate, V., Johnson, M. H., Osborne, T., & Csibra, G. (2009). Predictive motor activation during action observation in human infants. *Biology letters*, *5*(6), 769-772. doi:10.1098/rsbl.2009.0474

Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., & Aglioti, S. M. (2010). Simulating the future of actions in the human corticospinal system. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(11), 2511-2521. doi:10.1093/cercor/bhp292

Weiss, D. J., Wark, J. D., & Rosenbaum, D. A. (2007). Monkey see, monkey plan, monkey do: the end-state comfort effect in cotton-top tamarins (Saguinus oedipus). *Psychological science*, *18*(12), 1063-1068. doi:10.1111/j.1467-9280.2007.02026.x

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological bulletin*, *131*(3), 460-473. doi:10.1037/0033-2909.131.3.460

EXPERIMENTAL CONTRIBUTION

CHAPTER 4

One of the overarching objectives of the present thesis is to understand how the cognitive and behavioral characteristics of humans affect their ability to learn socially from observing their conspecifics, and to figure out what the consequences of these characteristics could be on the emergence of cultural traditions (i.e., the transmission and the stabilization among a population of behavioral patterns via social learning). However, such a prospect is of limited interest without testing the impact of individuals' cognitive and behavioral characteristics on the learning dynamics at stake at a population level. This is what we aimed to do in the last study presented in this thesis. The objective of the **Experimental Contribution Chapter 4** was to model behavioral constraints of individuals (such as the biomechanical constraints of the acting body) and to study their impact on i) their ability to learn socially from each other and on ii) the evolution of stable behavioral patterns among the population they belong to. To do so, we used a computer simulation procedure known as individual-based modeling.

Individual-based models (or also term Agent-based models) is a class of computational models for simulating the actions and interactions of autonomous agents (both individual or collective entities such as groups). The aim of these models is to assess the effects of these agents' interactions on the system as a whole. Individual-based models combine elements of game theory, complex systems, emergence, computational sociology, multi-agent systems, and evolutionary programming (Grimm and Railsback, 2005). They are used to explain the emergence of a variety of higher-order patterns, from network structures ranging from terrorist organizations to consumer behavior. Thus, individual-based models are perfectly adapted to simulate the simultaneous operations and interactions of multiple agents, in an attempt to re-create and predict the appearance of complex patterns at the population level, such as the emergence of behavioral traditions among a group through social transmission mechanisms.

The process is one of emergence from the lower (micro) level of systems to a higher (macro) level. The basic principles of individual-based models are that i) simple behavioral rules generate complex behavior, ii) the whole is greater than the sum of the parts. Individual agents are typically characterized as bounded rational agents, presumed to be acting in what they perceive as their own interests, such as reproduction, economic benefit, or social status, using heuristic learning or simple decision-making rules. Agents in individual-based models can thus experience learning, adaptation, and reproduction. In most of the cases, individual-based models are composed of i) numerous agents specified at various scales, ii) decision-making heuristics, iii) learning rules or adaptive processes, iv) an interaction topology; and v) a non-agent environment.

Accordingly, the aim of our model was to study, at a population level, the potential effect of various constraints (e.g., the size of the behavioral repertoire of an individual and the type of search space characterizing a task problem) on the ability of individuals to evolve behavioral traditions through the acquisition of faithful social learning. Even though they generally represent simplifications of reality, the use of evolutionary individual-based models is increasing in animal and human behavior studies. The reason is that they enable the identification of different selective pressures under varying ecological conditions, thus helping researchers to select the data needed to understand otherwise opaque phenomena (see also Acerbi et al., 2011).

Behavioral constraints and the evolution of faithful social learning

(2012, Current Zoology, 58(02), 307-318)

Alberto Acerbi[1], Pierre O. Jacquet[2,3], Claudio Tennie[4]

[1] Center for the Study of Cultural Evolution, University of Stockholm, Stockholm, Sweden

[2] Department of Psychology, University of Bologna, Bologna, Italy

[3] INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Center, ImpAct Team, Lyon, France

[4] Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract

Behavioral “traditions”, i.e. behavioral patterns that are acquired with the aid of social learning and that are relatively stable in a group, have been observed in several species. Recently, however, it has been questioned whether non-human social learning is faithful enough to stabilize those patterns. The observed stability could be interpreted as a result of various constraints that limit the number of possible alternative behaviors, rather than of the fidelity of transmission mechanisms. Those constraints can be roughly described as “internal”, such as mechanical (bodily) properties or cognitive limitations and predispositions, and “external”, such as ecological availability or pressures. Here we present an evolutionary individual-based model that explores the relationships between the evolution of faithful social learning and behavioral constraints, represented both by the size of the behavioral repertoire and by the “shape” of the search space of a given task. We show that the evolution of high-fidelity transmission mechanisms, when associated with costs (e.g. cognitive, biomechanical, energetic, etc.), is only likely if the potential behavioral repertoire of a species is large and if the search space does not provide information that can be exploited by individual learning. Moreover we show how stable behavioral patterns (“traditions”) can be achieved at the population level as an outcome of both high-fidelity and low-fidelity transmission mechanisms, given that the latter are coupled with a small behavioral repertoire or with a search space that provide substantial feedback. Finally, by introducing the possibility of environmental change, we show that intermediate rates of change favor the evolution of faithful social learning.

Keywords: animal social learning; cultural evolution; cultural transmission; copying fidelity; individual based modeling.

Introduction

Examples of behavioral "traditions", i.e. behaviors acquired with the aid of some forms of social learning, and which are relatively stable in groups, have been found in several species (Laland and Galef, 2009). The existence of these traditions has been usually considered as a strong indication of the presence of faithful social transmission mechanisms that guarantee both the successful diffusion and the stability of the behaviors involved (Huffman, 1996; Horner et al., 2006; Marino et al., 2007). In particular, since imitation - i.e. the high-fidelity copy of novel behaviors through the reproduction of action sequences of observed individuals (Call and Carpenter, 2002; Tennie et al., 2006; Whiten et al., 2009) - has often been viewed as the learning mechanism that best explained the emergence of human traditions (Boyd and Richerson, 1996; Tomasello et al., 1993; Tomasello, 1999), it is assumed that also non-human traditions are supported by similar imitative capacities (Claidière and Sperber, 2010).

However, it has been recently questioned whether non-human social learning is actually faithful enough to produce such stable behavioral patterns. For example, it has been shown that in experimental settings great apes tend to scarcely use imitation (Tennie et al., 2006; Tennie et al., 2009; Tennie et al., 2010). On a more theoretical side, Claidière and Sperber (2010) argued that the fidelity of social learning, as deduced by transmission chain studies in different species, may explain the propagation, but not the stability, of non-humans behavioral traditions.

Accordingly, researchers have begun to examine whether, and how, non-human animals, unequipped with faithful social learning capacities, could be able to develop behavioral traditions (Huber et al., 2009; Shea, 2009). It has been suggested that stable behavioral patterns could also result from transmission mechanisms less faithful than imitation, such as emulation, social and local enhancement, or even from trial-and-error learning (Caldwell and Millen, 2009; Franz and Matthews, 2010; Heyes, 1993; Laland and Hoppitt, 2003; Matthews et al., 2010; Whiten et al., 2003). Finally, it has been proposed that the observed stability could be the result of constraints that limit the number of possible alternative behaviors, more than the result of the fidelity of transmission mechanisms (Tennie et al., 2009; Claidière and Sperber, 2010).

In this paper we investigate the relationship between behavioral constraints and faithful social learning through an evolutionary individual-based model in which a hypothetical "species", first unequipped with high-fidelity copying mechanisms, may evolve them under different conditions. More specifically, we assumed that increasing the fidelity of social learning had some costs (e.g. cognitive, energetic, etc.) and also that the behavior of this species was variously constrained.

We introduced two kinds of constraints into our model. The first series of constraints limited the variety of individuals' behavioral repertoires. In real-life those constraints would translate into a set of "internal" factors, such as cognitive limitations (e.g. poor working memory capacities limit the number of behavioral sequences a species can plan or copy; van Leeuwen et al., 2009), cognitive biases (e.g. preference for certain classes of stimuli or certain types of demonstrators towards which one directs its behaviors; van de Waal et al., 2010), or bodily (biomechanical) architecture of acting individuals (e.g. limited degrees of freedom of effectors restrict the flexibility by which one can interact with external objects; Desmurget et al., 1995). In our model those constraints determined the number of the possible behavioral alternatives a species was provided with. Note that this indicates the distribution of *potential* behaviors, limiting the space in which the search for the optimal behavior is made, while the actual behaviors a population will show is a subset of those.

The second series of constraints pertained to the specific task one has to resolve and can be exemplified by a set of “external” factors (e.g. ecological) that shape the structure of the search space in which the candidate solutions takes place (Acerbi et al., 2011; see also Goldstone et al., 2008). For example, finding the ripest fruits on a tree is a very different problem with respect to choosing an edible fruit among different (perhaps including poisonous) fruits. In the former case an individual can try different fruits and, given adequate sensory and cognitive capacities, can choose to eat the sweetest ones; a strategy that is clearly not efficient in the latter situation. Here, we identified three distinct search spaces (see Figure 1), distinguished by the way payoffs were distributed among possible behaviors and, by consequence, by their tendency to enable individual search strategies. In the Methods section we describe the three spaces used in the model in detail and provide a real-life example for each.

The model we developed is individual-based (Grimm and Railsback, 2005), meaning that we simulated interactions at the level of single individuals, and evolutionary, i.e. an evolutionary algorithm (Holland, 1975) is used to optimize the behavior of individuals. The evolutionary algorithm acted on a variable that encoded the fidelity of social learning of each individual. Individuals that performed better resulted in proportionally more “offspring” than others. Even though they generally represent simplifications of reality, the use of evolutionary individual-based models is increasing in animal behavior studies. The reason is that they enable the identification of different selective pressures under varying ecological conditions, thus helping researchers to select the data needed to understand otherwise opaque phenomena (see also Acerbi and Nunn, 2011).

In the next section, we describe the implementation and the features of the model in detail, before presenting the results. We first investigated in which conditions a species, starting from completely unreliable social learning capacities, and thus relying upon individual learning only, was likely to evolve costly faithful social learning mechanisms. Secondly, we analyzed how population behavioral homogeneity (i.e. behavioral traditions) could be reached under different behavioral constraints, i.e. varying the size of the behavioral repertoire as well as the tasks' search spaces. Finally, we run other simulations allowing the possibility of environmental change, and we tested its effect on the evolution of faithful social learning. In the last section, we discuss the relevance of our results for the study of animal social learning and culture, limitations and possible extensions of our model, as well as some broad implications for modern human culture.

Methods

General description of the model

All simulations involved populations of individuals (N=100) that interacted in discrete time steps (until T=10000). At the beginning of the simulations each individual was assigned a behavior, randomly chosen among all possible behaviors characterizing its population. Populations varied with respect to the size of their behavioral repertoire (S): we distinguished three different experimental conditions, with populations disposing of a repertoire of 10, 100, or 200 possible behaviors.

At each time step, individuals interacted in pairs. Each individual (the *learner)* was paired with another individual (the *demonstrator*) randomly chosen among the ten individuals of the population with the highest payoffs (see below for how payoffs were calculated). Thus, each learner attempted to copy the behavior of its paired demonstrators. The accuracy of social learning depended on an individual characteristic, determined by the parameter α (fidelity of social learning). At the beginning of the simulations, α was initialized equal to zero for all individuals (making social learning completely unreliable for all individuals) and its value evolved through time.

Evolutionary dynamics resulted from a death-birth process in which newborns inherited the value of α from fittest individuals. Below we describe the details of the model's implementation.

Behavioral repertoire and search spaces

Experimental conditions varied with respect to the size of the population's behavioral repertoire (S=10, 100, and 200 possible alternative behaviors) as well as to how payoffs were distributed among possible behaviors, determining three different “search spaces”. Behavioral payoffs varied between 0 and 1, and only a single behavior, randomly selected, brought the maximum payoff to individuals in all spaces.

The smooth space

In *smooth spaces* (see Figure 1a), different behaviors laid on a payoff gradient, and the “closest” a behavior was to the optimal one, the higher its payoff. Payoffs were modeled as a Gaussian distribution (as in Boyd and Richerson, 1985 and Mesoudi and O'Brien, 2008). Smooth spaces represent tasks for which, even if an optimal solution exists, sub-optimal alternatives are similar in terms of payoff returns. Moreover, the existence of a payoff gradient provides a way to orient individual searches so that individual learning (e.g trial and error learning) can potentially be as effective as social learning (Acerbi et al., 2011). Even complex behaviors like chimpanzees' ant dipping have aspects that may be considered searches in smooth spaces (Humle and Matzusawa, 2002). An individual, for example, can repeatedly experiment with sticks of different length, self-evaluate the outcomes of different attempts, and then arrive at the measure that is most appropriate in a given situation.

The rugged space

Rugged spaces (see Figure 1b) represent “difficult” tasks for which only few good solutions exist. Contrary to smooth spaces, the structure of such tasks does not provide ways to orient individual searches. In our simulations rugged spaces were generated by assigning to every possible behavior a random payoff drawn from an exponential distribution with mean=1 (rescaled between 0 and 1), so that a single behavior led to the maximum payoff, while a restricted number of alternatives approximate it and a vast majority led to low payoffs. One real-life example of a task represented by a rugged search space could be foraging in a patchy, heterogeneous, environment (see e.g. Gil and Wolf, 1977). In such a situation, an individual can potentially try different sources of food, with only few of them being fruitful, without knowing in advance which one will be the richest. The knowledge of one source, furthermore, does not give information about the quality of the other sources present in the environment.

The peaked space

Finally, peaked spaces (see Figure 1c) represent even more difficult tasks, for which only a single behavior provides a payoff to individuals, distinguishing it from the two other search spaces described above. In Acerbi et al. (2011) we argued that many real-life tasks, especially in human culture, fit this description. One simple example is tying a knot: performing a behavior similar – but not equal – to the one requested to tying the knot does not produce a "less effective" knot, but in general does not produce any usable result. For this kind of task it is likely that any form of individual learning would be very ineffective, since there is nothing in the search space that could orient the search and there is only a single rewarding solution.

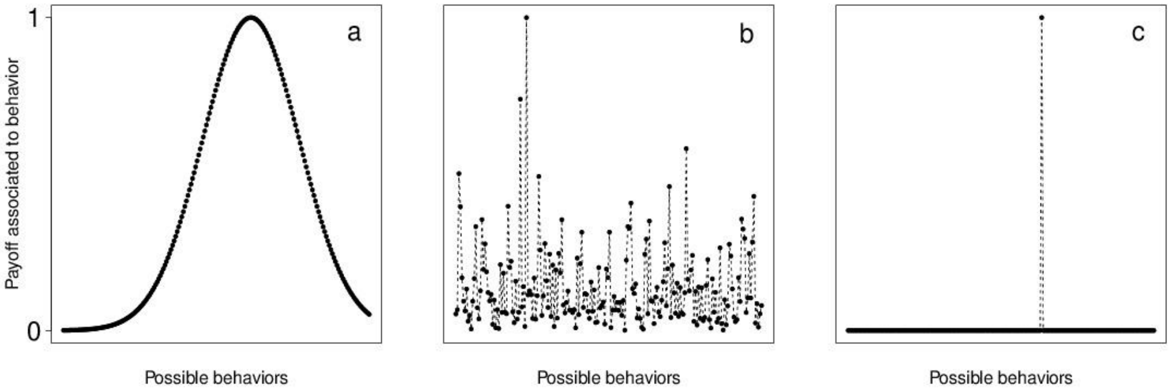


Figure 1. Search spaces used in the simulations. Schematic representation of the three payoff distributions used in the simulations determining the three different search spaces. (a) Smooth space. (b) Rugged space. (c) Peaked space. (See text for details).

The copying process

The outcome of the copying process depended on the observer's value of α. Figure 2 illustrates how the new behavior was picked up by the observer. Once the demonstrator was chosen, a new behavior was randomly selected in the search area included between $\pm \frac{S}{2}(1-\alpha)$ with respect to the demonstrator's behavior, and retained by the individual if its payoff was equal or higher with respect to the current payoff.

When α is close to 1 – such that the fidelity of the learner's copy is almost perfect – this expression is close to 0, meaning that individuals will assume a behavior closely approximating the demonstrator's behavior (with α =1 the copied behavior will be exactly the demonstrator's behavior, so, in this case, social transmission equates to replication). On the contrary, when α is close to 0, the expression is close to $\frac{S}{2}$, covering a large range of the behavioral repertoire. Since behaviors that decrease individual's payoff are discarded, α =0 can be considered cases of pure individual learning.

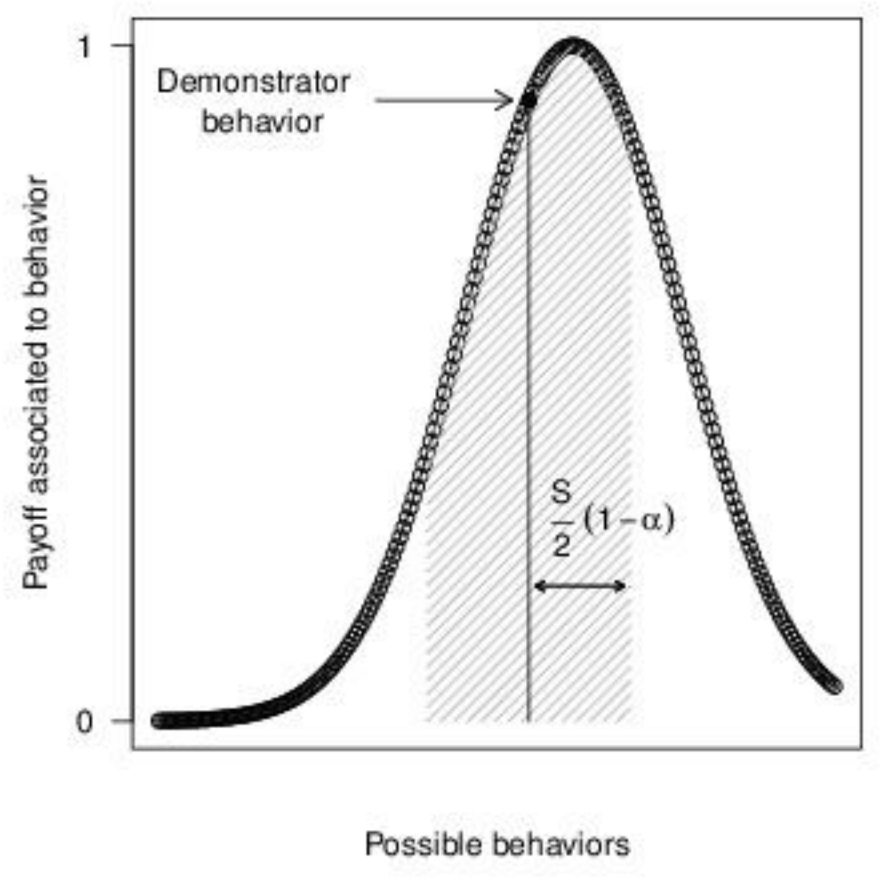


Figure 2. Schematic representation of the copying process (in a smooth search space). Given the demonstrator behavior, the learner will randomly pick up one behavior in the gray area. The size of the area is given by $\pm\frac{S}{2}(1-\alpha)$, where S is the size of the population's behavioral repertoire and α represents the learner's fidelity of social learning. (See text for details).

Average payoff and evolutionary algorithm

A basic assumption of our model was that faithful social learning has some cost, and this cost modulated the payoff an individual received from performing a behavior. Individual payoff was hence determined both by the behavior performed and by the fidelity of social learning represented by their value of α .

In more detail, the payoff of individuals (P_i), at each time step, was equal to:

$$P_i = P_b - \alpha C$$

that is, the payoff obtained by the behavior performed (P_b), minus the value of α multiplied for a factor *C:*

We varied the value of *C* from 0 (no cost) to 0.5 (highest cost), with steps of 0.1, representing alternative situations in which using (and evolving) faithful social learning could be more or less costly.

The *average* individual payoff, used to select individuals for reproduction in the evolutionary algorithm, was simply the sum of all payoffs an individual had had in the course of its life, averaged for the number of time steps it was alive.

Individuals were selected for reproduction according to their average payoff. At each time step one individual, randomly chosen among the entire population, was replaced by a "newborn". The newborn individual inherited the value of α from another individual that was randomly chosen among the ten individuals with the highest average payoffs. The behavior of the newborn was initialized randomly. In other words, the fidelity of social learning, and not the behavior *per se*, was genetically inherited and, hence, subject to evolutionary pressures.

Finally, with a small probability of mutation (μ =0.05), the inherited value of α was randomly reinitialized with a value comprised between 0 and 1.

Simulation procedures

In a first set of simulations we studied three different sizes of behavioral repertoires (S=10, 100, and 200) for each payoff distribution (smooth space, rugged space, and peaked space). For each condition, we varied the cost factor of fidelity (C=0, 0.1, 0.2, 0.3, 0.4, and 0.5) and we ran 100 simulations for every value of C, recording the average value of fidelity evolved.

We then analyzed how, in peaked search spaces, the interaction between the fidelity of social learning and the size of the behavioral repertoire impacted the populations' behavioral diversity, namely, the number of behavioral patterns present in a population. To calculate behavioral diversity we used Simpson's diversity index. Simpson's diversity index was developed mainly to assess ecological diversity, taking into account both the number and relative abundance of species present in a given environment (Simpson, 1949). Recently it has been used to assess behavioral diversity in cultural evolutionary models (Kandler and Laland, 2009; Enquist et al., 2010). According to this index the diversity of a population can be represented as:

$$D=1-\sum_{i}\chi_{(i)}^{2}$$

where $\chi_{(i)}$ is the frequency of the *i* variant in the population. The value of *D* tends towards 1 as the behavioral diversity of a population increases, and is equal to 0 when all individuals share the same behavioral variant.

We also measured directly the number of existing behaviors at the end of simulations, comparing the effect of the three different search spaces, keeping the other parameters constant (S=200, C=0.2).

In a second set of simulations, we added the possibility of environmental changes. An additional parameter (p_c=0.001, 0.01, and 0.1) determined at each time step the probability of the payoff distribution to be fully reinitialized. Note that reinitialization of payoffs did not change the structure of the search space but it changed the distribution of payoffs among the behaviors. In smooth spaces as well as in peaked spaces, this involved the "shifting" of the optimal behavior in a different position, and in rugged spaces the re-assignation of a random payoff to every possible behavior drawn from the exponential distribution described in subsection 2.2.2. Notice that when p_c=0 this condition reduces to the basic simulation.

Symbol	**Short description**	**Values**
N	Population size	100
T	Number of time steps	10000
S	Size of the behavioral repertoire	**10, 100, 200**
C	Cost factor of fidelity	**0, 0.1. 0.2, 0.3, 0.4, 0.5**
μ	Mutation rate	0.05
p_c	Probability of environmental change	**0, 0.001, 0.01, 0.1**

Table 1. Main parameters and their value used in the model. Bold typeface values are values varied in different experimental conditions.

Results

Faithful social learning evolves with large behavioral repertoire and in peaked spaces

The results of our simulations showed that, in stable environments, costly faithful social learning evolved only if two conditions were simultaneously met: the populations had a large behavioral repertoire, and the task structure was a peaked space (see Figure 3).

When faithful social learning was cost-free (C=0), all populations converged towards high average values of α, showing that high-fidelity transmission mechanisms proved advantageous to individuals in all conditions. However, when copying mechanisms involved costs that impacted on the individual's payoff (C>0), these costs were only worthwhile in situations where an individual search was ineffective. This occurred in peaked search spaces, and with populations characterized by a sufficiently large behavioral repertoire (see Figure 3c).

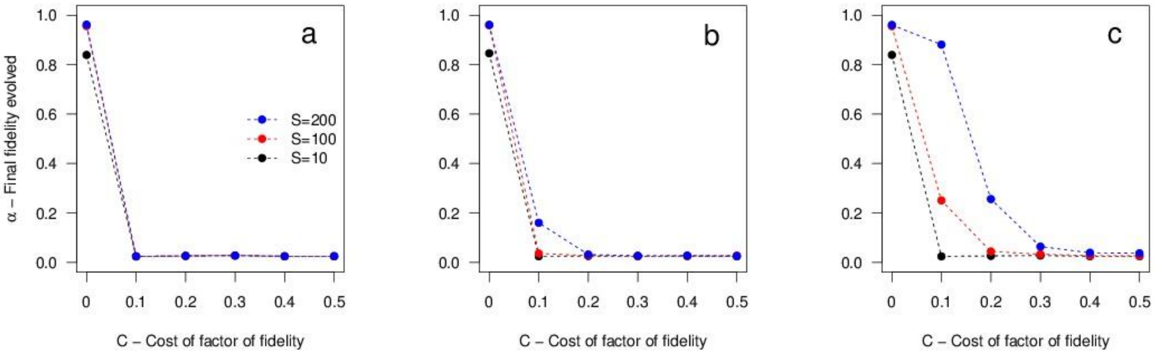


Figure 3. **Average fidelity evolved at the end of simulations versus cost factor of fidelity**. (a) Smooth space. (b) Rugged space. (c) Peaked space. Different lines colors in the three conditions represent different sizes of the behavioral repertoire: blue line: S=200; red line: S=100, black line: S=10. (Each data point is an average on 100 runs).

Behavioral homogeneity results from both high-fidelity social learning mechanisms and low-fidelity mechanisms

Figure 4 depicts the evolutionary trajectories of typical runs in the peaked space condition, for populations having different sizes of behavioral repertoire and for three different values of C. We have chosen to analyze in detail the peaked space condition because in this condition faithful social learning evolves also when is costly (see results in Figure 3). Each point in the plots represents the "position" of a population with respect to its behavioral diversity (x-axes) and its average fidelity of social learning (y-axes), sampled at an interval of 100 time steps, during each run. Populations that are in the left part of the graphs are behaviorally homogeneous populations and populations that are on the right side are behaviorally diverse. With respect to y-axes, populations that are in the lower part of the graphs lack hi-fidelity social learning abilities while population in the upper part posses them.

Populations always "started" in the bottom right corner of the graphs, i.e. they were diverse (at the beginning of the simulations behaviors were randomly initialized) and individuals also did not possess hi-fidelity social learning abilities (α was initialized at 0 for all individuals). When faithful social learning was cost-free (C=0, Figure 4 left), populations, irrespective of their behavioral repertoire's size, "moved" towards the high left corner of the plot during the simulation run, i.e. towards behavioral homogeneity and faithful social learning. It is worth noting that populations with a small behavioral repertoire (black line) could move to the left area of the plot (i.e. towards behavioral homogeneity) without individuals being required to increase their social learning ability (this happened only in later stages of the simulation). On the contrary, populations with larger behavioral repertoire (blue and red lines) were required to increase the faithfulness of social learning ("moving up in the plot") in order to move towards behavioral homogeneity.

For intermediate costs of faithful social learning (C=0.1, Figure 4 center), the evolutionary trajectories of populations with large behavioral repertoire were similar, while the population with small behavioral repertoires reached homogeneity without developing faithful but costly social learning. Finally, when faithful social learning was even more costly (C=0.2, Figure 4 right), even populations with large behavioral repertoires did not evolve it, and their behavioral diversity remained high.

In sum, while behaviorally diverse populations (right part of the plots) were the outcome of a large behavioral repertoire coupled with low-fidelity social learning mechanisms (Figure 4 right), behavioral homogeneity (i.e. low diversity, left part of the plots) could be the product either of faithful social learning (Figure 4 left) or of low-fidelity social learning, provided that the behavioral repertoire was small (Figure 4 center, black line).

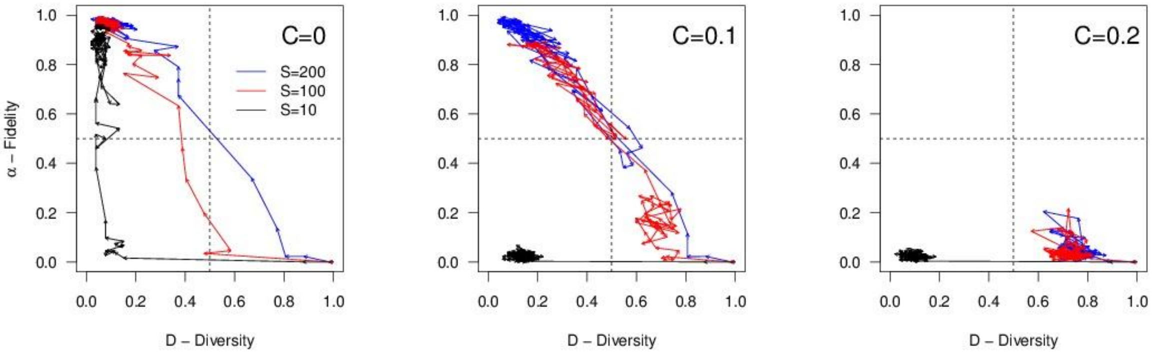


Figure 4. **Evolutionary trajectories of populations.** Evolutionary trajectories of populations in respect to behavioral diversity (x-axes) and average fidelity of social learning (y-axes). Each point in the plot represent the "position" of a population at a given stage of the evolution (sampled every 100 time steps in a simulation run). Different lines colors represent different sizes of the behavioral repertoire: blue line: S=200; red line: S=100, black line: S=10. The cost factor of social learning varies in the three panels: from left to right C=0, C=0.1, C=0.2.

However, the size of the potential behavioral repertoire was not the only factor that influenced the final behavioral diversity of a population. We analyzed simulation runs with the same behavioral repertoire size (S=200) and the same cost factor for social learning (C=0.2) and we measured the number of behaviors present at the end of the simulations for the three different search spaces (Figure 5). In smooth and rugged spaces, where faithful social learning did not evolve (see results in Figure 3 a and b), the populations showed approximately 20 different behaviors. In peaked search paces, however, the final number of behavior in absence of faithful social learning was higher, and the same number of behaviors was reached only when faithful social learning evolved.

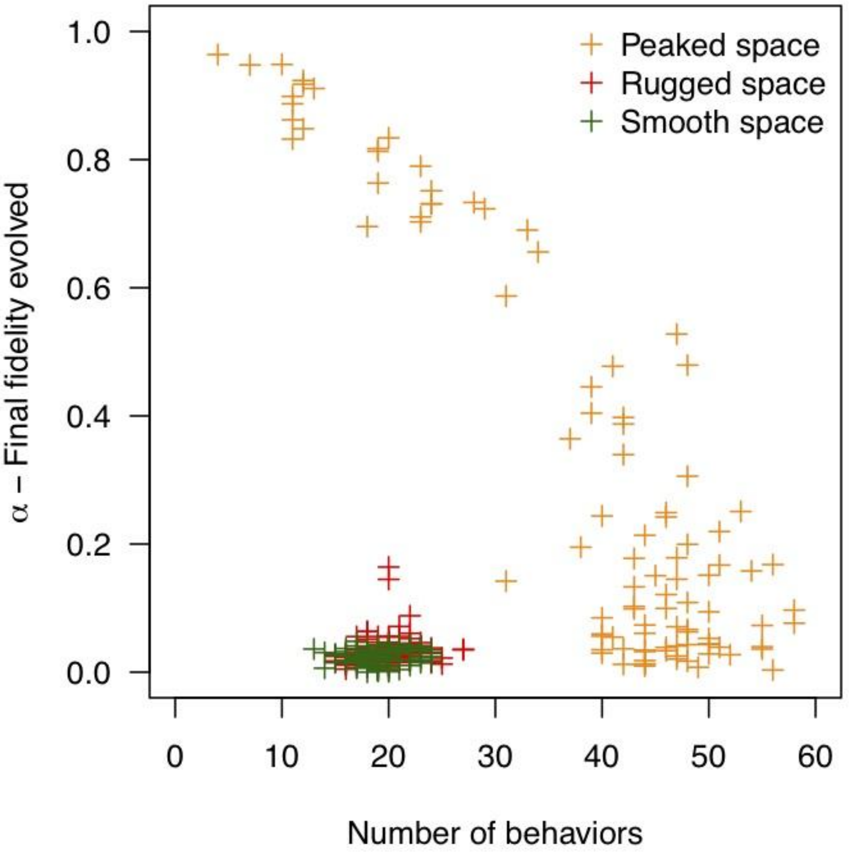


Figure 5. Final number of behaviors present in populations in the three different search spaces. Each point of the plot represents the results of a simulation in respect to the final number of behaviors present in the population (x-axes) and the average fidelity of social learning evolved (y-axes), keeping fixed the factor cost of social learning (C=0.2) and the size of the behavioral repertoire (S=200). Orange: peaked space, red: rugged space, green: smooth space. (For each space 100 simulations were run).

Intermediate rates of environmental change favor the evolution of faithful social learning

Finally, we analyzed the effect of environmental variation on the evolution of faithful social learning, running additional simulations for populations with a large behavioral repertoire (S=200). Populations with a large behavioral repertoire were specifically targeted as the above described results showed that in these conditions faithful social learning was more likely to evolve when the environment was fixed.

In smooth search spaces, environmental variation had no effect on the evolution of faithful social learning (Figure 6a), and populations remained composed of individuals with poor copying abilities, even when the environment was variable. For rugged and peaked search spaces (Figure 6b and 6c), we found instead that the evolution of faithful social learning was favored for intermediate rates of environmental variation. In fact, when the rate of environmental variation was too high (p_c=0.1) the average values of fidelity evolved were similar to the condition in which the environment was stable (p_c=0).

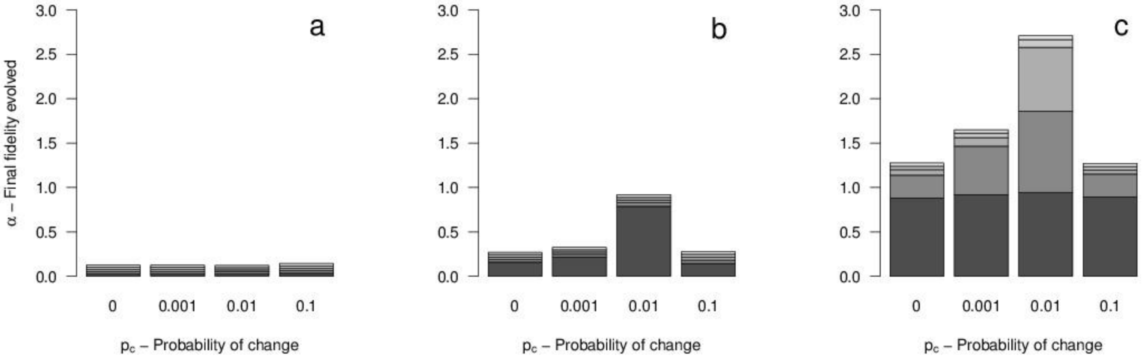


Figure 6. Stacked bar plot of the average fidelity evolved with different probabilities of environmental change for population with S=200 (size of the behavioral repertoire). The different colors in the bars represent different values of C (factor cost of faithful social learning) from C=0.1 (darker) to C=0.5 (lighter). We did not take into account C=0 because in this condition faithful social learning always evolved for p_c=0 (see Figure 3). (a) Smooth space. (b) Rugged space. (c) Peaked space. (Each data is an average on 100 runs).

Discussion

General discussion of the results

The present individual-based model examined the relationship between behavioral constraints and the evolution of faithful social learning. Constraints varied according to the size of the behavioral repertoire of populations (10 vs. 100 vs. 200 possible alternative behaviors) and according to the intrinsic structure of the search space characterizing the task problem (*smooth space* vs. *rugged space* vs. *peaked space*). We firstly analyzed, by varying the cost for individuals to use faithful transmission mechanisms (from *null* to *high* cost), how and in which type of search spaces populations with different sizes of behavioral repertoire would take advantage of such faithful social learning. We also took into account the effects of fidelity of social learning and behavioral constraints on the behavioral diversity at population level. We investigated whether populations unequipped with high-fidelity transmission mechanisms were prone to develop and stabilize novel behavioral patterns in a manner outwardly similar to populations equipped with high-fidelity transmission mechanisms. Finally, the effect of the rate of environmental change in which populations evolved (from *no* to *fast* environmental change) was studied.

Three main results emerged. First, in stable environments, costly faithful social learning evolved only in populations with large behavioral repertoires, and particularly in peaked search spaces. Second, the convergence towards behavioral homogeneity resulted from high-fidelity social learning mechanisms but also from low-fidelity mechanisms, when they were associated with a small behavioral repertoire or with smooth and rugged search spaces.

Third, intermediate rates of environmental change favored the evolution of faithful social learning.

The evolution of faithful social learning, when costly, strongly depended on behavioral constraints. According to our results, we should expect to find, in real-life, faithful social learning in conditions in which a species, or a group, has many behavioral alternatives (a large *S* in our model) and, at the same time, in which the task at hand does not provide any structure useful to orient the individual's search. In particular these conditions were met in peaked spaces. Tasks characterized by this search space had two interesting features. First, only a very narrow number of behaviors – in our model, only one – led to success (i.e. the payoff achievable with sub-optimal behaviors is zero, differently from other spaces). Secondly, performing behaviors other than the single successful solution did not provide any feedback that individuals could use to estimate the optimality of a given behavior. This result confirms and enriches our previous findings (Acerbi et al., 2011) where we showed that, for tasks whose search structure could be modeled as a peaked space, imitation – i.e. a specific instance of high-fidelity social learning mechanism – was more effective than emulation and individual learning.

We also showed that the convergence of a population towards behavioral homogeneity could result, as expected, from high-fidelity social learning mechanisms, but also from low-fidelity social learning mechanisms. An analysis of simulations for peaked search spaces demonstrated that a population with a small behavioral repertoire could become behaviorally homogeneous without developing high-fidelity social learning mechanisms. Additionally, our results also showed that when the number of potential behaviors was large, the search structure had an impact on the number of behaviors actually present in the population. In particular, relative homogeneity in absence of high-fidelity social learning mechanisms was obtained for smooth and rugged search spaces, but not for peaked spaces.

This observation is of importance since the emergence of behavioral homogeneity in wild populations is often presumed to be a sign of faithful social learning (Huber et al., 2009). While this could certainly be the case – in the simulations presented here faithful social learning does indeed produce behavioral homogeneity – our model provides an alternative explanation. This explanation is based on the existence of behavioral constraints, may they be due to physical and/or cognitive limitations, or ecological factors (shaping the search space of a given task). It has been shown that behavioral constraints can lead to the re-appearance of presumed cultural behaviors in naïve captive individuals (Huffman and Hirata, 2004; Tennie et al., 2008; see also Masi, 2011). With regard to ecological influences, it has long been suggested that these may help explain the distribution of several behaviors across populations (Humle and Matzusawa, 2002; though see Schöning et al., 2008; Möbius et al., 2008). For a behavior presumed to be a product of faithful social transmission, one has to check whether its diffusion among the population is accounted for by such alternative possibilities (see also Laland and Janik, 2006; Tennie et al., 2009). Of course, as nearly always in modeling, our model represents an ideally simplified situation. However, one could imagine having an estimation of the possible alternative behaviors a species is likely to use (see e.g. Changizi, 2003), as well as an estimation of the search structure of a specific task (for example the distribution of resources in a specific environment and their energetic/caloric contribution). These data can then be used to parameterize the model. In this way one could obtain more realistic results that could be used as a guide to analyze whether, in a specific situation, a given population is likely to make use of social learning.

Finally, by manipulating the probability of environmental change, we showed that intermediate rates of environmental change favored the evolution of faithful social learning. Importantly, with moderate rates of environmental change, costly faithful social learning evolved not only in peaked spaces, but also in rugged search spaces. This is consistent with the idea that the three search spaces we modeled represent three different levels of "difficulty" (see below however for how we intend the meaning of "difficulty" here), with smooth spaces representing "easy" tasks, followed by rugged spaces, and then by peaked spaces as the most difficult ones. More generally, this result is coherent with the broad consensus that the evolution of social learning is more likely to occur for an intermediate rate of environmental change than for no change – where genetic evolution is favored - or fast change – where individual learning is favored (see e.g. Henrich and McElreath, 2003; Wakano et al., 2004; Aoki et al., 2005).

Related literature and possible extensions of the model

The results of our model are, in general, consistent with the "costly information hypothesis" (Boyd & Richerson, 1985), according to which social learning is favored when acquiring information individually is costly or inaccurate (see e.g. Rogers, 1988; Boyd & Richerson, 1995; Wakano et al., 2004; Aoki et al., 2005). We aimed to illustrate how this trade-off between social and individual learning could be realized in a scenario analogous to many real-life situations, focusing on the notion of behavioral constraints, and we believe that this illustration may be of some use for field biologists and comparative psychologists who study social learning and cultural evolution.

The role of the variation of search spaces, or adaptive landscapes, have been considered in previous models of cultural evolution (Boyd & Richerson, 1992; Mesoudi, 2008). These models show how multimodal adaptive landscapes – i.e. search spaces with more than one peak – favor social learning, contrary to unimodal adaptive landscapes, where individual learning is favored. Our results add to these previous finding by showing that also in unimodal adaptive landscapes social learning may be favored, as long as the search space does not provide information that can be used to orient individual learning (our peaked space condition).

Previously, other computational models (Hinton & Nowlan 1987) had shown that problems analogous to tasks represented by the peaked space could be solved through a combination of individual learning and genetic evolution. Since we did not consider genetic evolution (i.e. our evolutionary algorithm acted on the accuracy of social learning, and not on the actual behavior), our model is unable to address this question, though we obtained the same qualitative result with respect to the poor performance of individual learning alone. In a later development of Hinton and Nowlan's model (Best, 1999), the possibility of social learning was added, and it was shown that, indeed, the combination of social learning and genetic evolution improved the performance compared to the combination of individual learning and genetic evolution. In Best's model, however, social learning was cost-free and no changes in the search space or in the size of the behavioral repertoire were taken into account.

As with many models, we concentrated here on few parameters that we thought of fundamental importance for our study, namely the cost of acquiring faithfully social information, the size of the behavioral repertoire, the different search spaces, and, as a final check of the validity of our model, the extent of environmental variation. Interesting developments could consist in examining the effects of other factors on the results here reported. For example, we used a basic evolutionary algorithm, mainly intended as a proof-of-concept tool, keeping a fixed – and high – selection pressure, a simple implementation of the mutation – α was reinitialized every time a mutation occurs –, and we did not consider the effect of population size on evolutionary dynamics. We had run some exploratory simulations to test the effects of the variation of these parameters (not reported in the results section). Varying population size (N=200; N=500) and changing the way mutations were implemented (α each time modified by a value randomly selected between -0.1 and +0.1) did not seem to change qualitatively our main results. Selection pressure, however, had some impact on the results. Interestingly, *less* selection pressure ("reproducing" individuals randomly chosen among the twenty, or fifty, individuals with the highest average payoffs) favored the evolution of social learning when costs were high, at least in peaked and rugged search spaces, and for large behavioral repertoires (S=100; S=200). We interpret this result as meaning that, with high selection pressures, "lucky" individuals that found optimal behaviors without using costly social learning were highly favored by the evolutionary algorithm, making populations of social learners unstable. The interactions between population size and selection pressure are anyhow inherently complex, and we plan to explore their effect on the evolution of faithful social learning in scenarios like ours in future works.

We also assumed that individuals were randomly paired in their interactions, a part from the fact that only individuals with proportionally high payoffs were targeted as possible demonstrators. Starting from the same set-up, one could certainly include more realistic rules of interactions, considering for example individuals being in different ways selective in their decisions about when and from who to copy (for the importance of these and other social learning "strategies" see Laland, 2004; Rendell et al. 2011), or explicitly consider a spatial dimension in the model, with individuals having different movement "rules" and interactions constrained by physical proximity.

As a final remark, we initialized our populations with random behaviors, chosen among all the possible alternatives of their potential behavioral repertoire. This is possibly an unrealistic situation (real populations do not show highly diverse – and certainly not random – behaviors) but we believe such a simplification to be useful as a "starting point" for the evolutionary algorithm. Again, future work could analyze how behavioral constraints impact on the evolution of social learning, starting from homogeneous populations that behave sub-optimally or already optimally, in which case social learning would be necessary to maintain the correct behavior through time.

General considerations and implications for modern human culture

We conclude with some general considerations derived from our results. In social learning research, the complexity of a task is often considered suggestive of the presence of social learning – with "easy" tasks being solved with individual learning and "difficult" tasks needing social transmission (see also Acerbi et al., 2011). Especially in laboratory tasks, experimenters try to propose "difficult" tasks to animals to encourage the use of social information to solve them (Day et al., 2003, Baron et al., 1996, Laland, 2004, Tennie et al., 2009). While this is probably a good rule of thumb, our model suggests that what makes faithful social learning useful is not the difficulty of a task *per se* (see also Tennie and

Hedwig, 2009) but the fact that relatively unconstrained behavioral alternatives are potentially involved in that task (or, if you prefer, a task is “difficult” when this happens). A spider's web may or may not be less complex than potato washing, what is different is that, in the former case, natural selection, working on the genetic level, highly constrained the behavioral repertoire, narrowing down the possible alternatives.

Within this perspective, even very complex human “cultural” behaviors may be a result of a combination of a genetically/ecologically narrowed behavioral repertoire, constrained search spaces, and some form of social learning (Sperber, 1996). One might consider, for example, cookery traditions. The impressive variability of foods consumed in different cultures is the outcome of various forms of cultural transmission (between and within societies) that nonetheless act on a “constrained space”: ecologically constrained (local availability of products), genetically constrained (only some products are edible; some taste preferences are at least partly innate, Rozin, 1990), and technologically constrained (many products have to be processed in a specific way to became edible; the technologies available in a group limit the choice of processing food techniques). On the other side, some cultural behaviors are relatively less constrained. Many fashions and fads, for example, result from pure transmission processes (see e.g. Bentley et al., 2007): the fashion of, say, “wearing green” one year but not the next has not much to do with behavioral constraints as we intended them in this paper. Analogously, if we take into consideration highly complex technological tasks, products of human cumulative culture (Richerson and Boyd, 2005), behavioral constraints become less and less important. Building a kayak – or an airplane – is certainly subject to constraints (all in all airplanes need to fly and kayaks need to float) but their guidance is so loose that only high fidelity copying mechanisms can allow an individual to acquire the necessary skills to produce them. Humans, nevertheless, also excel in a parallel strategy to solve those problems: cultural “epistemic engineering” (Sterelny 2003) is, according to the view presented here, a matter of narrowing the alternative solutions to a problem, and artificially build highly informative search spaces so as to reduce the need of costly social learning.

In conclusion, we believe that an explicit attention towards what is learned, and towards the potential alternatives and constraints, may enrich the theoretical toolbox of social learning modeling, and possibly our understanding of humans and other species' culture.

References

Acerbi A, Nunn CL, 2011. Predation and the phasing of sleep: an evolutionary individual-based model. Animal Behaviour 81: 801-811.

Acerbi A, Tennie C, Nunn CL, 2011. Modeling imitation and emulation in constrained search space. Learning & Behavior, 39: 104-114.

Aoki K, Wakano JY, Feldman MW, 2005. The emergence of social learning in a temporally changing environment: A theoretical model. Current Anthropology, 46: 334-340.

Baron RS, Vandello JA, Brunsman B, 1996. The forgotten variable in conformity research: The impact of task importance on social influence. Journal of Personality & Social Psychology, 71: 915-927.

Bentley RA, Lipo CP, Herzog HA, Hahn M, 2007. Regular rates of popular culture change reflect random copying. Evolution and Human Behavior, 28: 151–158.

Best ML, 1999, How culture can guide evolution: an inquiry into gene/meme enhancing and opposition. Adaptive Behavior, 7: 289-306.

Boyd R, Richerson PJ, 1985. Culture and the evolutionary process. Chicago, IL: University of Chicago Press.

Boyd R, Richerson PJ, 1992. How microevolutionary processes give rise to history. In MK Nitecki & DV Nitecki (Eds.), History and evolution (pp. 179-209). New York, NY: SUNY Press.

Boyd R, Richerson PJ, 1995. Why does culture increase human adaptability? Ethology and Sociobiology, 16: 125-143.

Boyd R, Richerson PJ, 1996. Why culture is common but cultural evolution is rare. Proceedings of the British Academy, 88: 77-93.

Caldwell CA, Millen AE, 2009. Social learning mechanisms and cumulative cultural evolution. Is imitation necessary? Psychological Science, 20: 1478-1483.

Call J, Carpenter M, 2002. Three sources of information in social learning. In K Dautenhahn & CL Nehaniv (Eds.), Imitation in Animals and Artifacts (pp. 211-228). Cambridge, MA: MIT Press.

Changizi MA, 2003. Relationship between number of muscles, behavioral repertoire size, and encephalization in mammals. Journal of Theoretical Biology. 220: 157-68.

Claidière N, Sperber D, 2010. Imitation explains the propagation, not the stability of animal culture. Proceedings of the Royal Society of London. Series B, Biological Sciences, 277: 651-659.

Day R, Coe RL, Kendal JR, Laland KN, 2003. Neophilia, innovation and social learning: A study of intergeneric differences in Callitrichid monkeys. Animal Behaviour, 65: 559-571.

Desmurget M, Prablanc C, Rossetti Y, Arzi M, Paulignan Y, Urquizar C, Mignot J-C, 1995. Postural and synergic control of three-dimensional movements of reaching and grasping. Journal of Neurophysiology, 74: 905-910.

Enquist M, Strimling P, Eriksson K, Laland K, Sjöstrand J, 2010. One cultural parent makes no culture. Animal Behaviour, 79: 1353-1362.

Franz M, Matthews LJ, 2010. Social enhancement can create adaptive, arbitrary and maladaptive cultural traditions. Proceedings of the Royal Society of London. Series B, Biological Sciences, 277: 3363-3372.

Gill FB, Wolf LL, 1977. Nonrandom foraging by sunbirds in a patchy environment. Ecology, 58: 1284-1296.

Goldstone RL, Roberts ME, Mason W, Gureckis T, 2008. Collective search in concrete and abstract spaces. In Kugler T, Smith C, Connelly T, Son Y-J (Eds.), Decision modeling and behavior in uncertain and complex environments (pp. 277-308). New York, NY: Springer.

Grimm V, Railsback SF, 2005. Individual-based Modeling and Ecology. Princeton, NJ: Princeton University Press.

Henrich J, McElreath R, 2003. The Evolution of Cultural Evolution. Evolutionary Anthropology, 12: 123-135.

Heyes C, 1993. Imitation, culture and cognition. Animal Behaviour, 46: 999-1010.

Hinton GE, Nowlan SJ, 1987. How learning can guide evolution. Complex Systems, 1: 495-502.

Holland JH, 1975. Adaptation in Natural and Artificial Systems. Ann Arbor, MI: University of Michigan Press.

Horner V, Whiten A, Flynn E, de Waal FB, 2006. Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. Proceedings of the National Academy of Sciences USA, 103: 13878-13883.

Huber L, Range F, Voelkl B, Szucsich A, Virányi Z, Miklosi A, 2009. The evolution of imitation: what do the capacities of non-human animals tell us about the mechanisms of imitation? Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364: 2299-2309.

Huffman MA, 1996. The study of nonhuman primate culture in Japan. In Heyes CM, Galef BG (Eds). Social learning in animals: the roots of culture (pp-267-289). San Diego, CA: Academic Press.

Huffman MA, Hirata S, 2004. An experimental study of leaf swallowing in captive chimpanzees: insights into the origins of self-medicative behavior and the role of social learning. Primates, 45: 113-118.

Humle T, Matsuzawa T, 2002. Ant dipping among chimpanzees of Bossou, Guinea and some comparisons with other sites. American Journal of Primatology, 58:133–148.

Kandler A, Laland KN, 2009. An investigation of the relationship between innovation and cultural diversity. Theoretical Population Biology, 76: 59-67.

Laland KN, 2004. Social learning strategies. Learning & Behavior, 32: 4-14.

Laland KN, Hoppitt W, 2003. Do animals have culture? Evolutionary Anthropology, 12: 150-159.

Laland KN, Janik VM, 2006. The animal cultures debate. Trends in Ecology and Evolution, 21: 542-547.

Laland KN, Galef BG, 2009. The question of animal culture. Cambridge, MA: Harvard University Press.

van Leeuwen ML, van Baaren RB, Martin D, Dijksterhuis A, Bekkering H, 2009. Executive functioning and imitation: Increasing working memory load facilitates behavioural imitation. Neuropsychologia. 47: 3265-3270.

Marino L, 2007. Cetaceans have complex brain for complex cognition. PLoS Biology, 5: e139.

Masi S, 2011. Differences in gorilla nettle-feeding between captivity and the wild: local traditions, species typical behaviors or merely the result of nutritional deficiencies? Animal Cognition. 14:921-925.

Matthews LJ, Paukner A, Suomi SJ, 2010. Can traditions emerge from the interaction of stimulus enhancement and reinforcement learning? An experimental model. American Anthropologist, 112: 257-269.

Mesoudi A, 2008. An experimental simulation of the “copy–successful–individuals” cultural learning strategy: adaptive landscapes, producer-scroungers dynamics, and informational access costs. Evolution and Human Behavior, 29: 350-363.

Mesoudi, A, O’Brien MJ, 2008. The cultural transmission of great basin projectile-point technology I: an experimental simulation. American Antiquity, 73: 3–28.

Möbius Y, Boesch C, Koops K, Matsuzawa T, Humle T, 2008. Cultural differences in army ant predation by West African chimpanzees? A comparative study of microecological variables. Animal Behaviour, 76: 37-45.

Rendell LE, Fogarty L, Hoppitt, WJE, Morgan TJH, Webster MM, Laland KN, 2011. Cognitive culture: Theoretical and empirical insights into social learning strategies. Trends in Cognitive Sciences, 15: 68-76.

Richerson PJ, Boyd R, 2005. Not by genes alone: How culture transformed human evolution. Chicago, IL: University of Chicago Press.

Rogers, AR, 1988. Does biology constrain culture? American Anthropologist 90: 819-831.

Rozin P, 1990. Development in the food domain. Developmental Psychology, 26: 555–562.

Schöning C, Humle T, Möbius Y, McGrew WC, 2008. The nature of culture: technological variation in chimpanzee predation on army ants revisited. Journal of Human Evolution, 55: 48-59.

Shea N, 2009. Imitation as an inheritance system. Philosophical Transactions of the Royal

Society of London. Series B, Biological Sciences, 364: 2429-2443.

Simpson EH, 1949. Measurement of diversity. Nature, 163: 688.

Sperber D, 1996. Explaining culture. A naturalistic approach. Oxford: Blackwell.

Sterelny K, 2003. Though in an hostile world. The evolution of human cognition. Oxford: Blackwell.

Tennie C, Call J, Tomasello M, 2006. Push or pull: emulation versus imitation in great apes and human children. Ethology, 112: 1159-1169.

Tennie C, Hedwig D, Call J, Tomasello M, 2008. An experimental study of nettle feeding in captive gorillas. American Journal of Primatology, 70: 584-93.

Tennie C, Hedwig D, 2009. How latent solution experiments can help to study differences between human culture and primate traditions. In Potocki E, Krasinski J (Eds), Primatology: Theories, Methods and Research (pp 95-112). New York, NY: Nova Publishers.

Tennie C, Call J, Tomasello M, 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364: 2405-2415.

Tennie C, Call J, Tomasello M, 2010. Evidence for emulation in chimpanzees in social settings using the floating peanut task. PloS ONE, 5: e10544.

Tomasello M, 2009. The cultural origins of human cognition. Cambridge, MA: Harvard University Press.

Tomasello M, Kruger AC, Ratner HH, 1993. Cultural learning. Behavioral and Brain Sciences, 16: 495-552.

van de Waal E, Renevey N, Favre CM, Bshary R, 2010. Selective attention to philopatric models causes directed social learning in wild vervet monkeys. Proceedings of the Royal Society of London. Series B, Biological Sciences, 277: 2105-2111.

Wakano, JY, Aoki, K., Feldman MW, 2004, Evolution of social learning: a mathematical analysis. Theoretical Population Biology, 66: 249-258.

Whiten A, Horner V, Marshall-Pescini S, 2003. Cultural panthropology. Evolutionary Anthropology, 12: 92-105.

Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM, 2009. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. Philosophical Transactions of the Royal Society. Series B, Biological Sciences, 364: 2417–2428.

GENERAL DISCUSSION AND PERSPECTIVES

This thesis had several objectives. The experiment described in Chapter One focused on a perceptual aspect of action understanding, and aimed to investigate the role of the inferior frontal cortex (IFC), the anterior intraparietal region (AIP) (thought to form part of the Action Observation Network – AON) and the primary somatosensory (S1) cortex in the hierarchical decoding of observed actions. The experiment described in Chapter Two investigated the contribution of prior information and sensorimotor constraints to action understanding and prediction, while Chapter Three's experiment was designed to probe whether the interaction between these two variables modulates corticospinal excitability (CSE) during action prediction. Finally, in Chapter Four we attempted to model at a population level the impact of these behavioral constraints on the emergence and maintenance of behavioral traditions acquired by means of social transmission mechanisms. Even though a huge amount of literature has been dedicated to the study of action understanding and its role in social learning, the nature of the relationships between these two issues is still a matter of debate. Overall, the experimental studies presented in this thesis invrease our understanding of these relationships. In particular, within this domain there is a lack of evidence regarding whether, and how, high-level (e.g., acquired from probabilistic exposure) and low-level (e.g., derived from the estimation of biomechanical costs engaged in the observed action) prior expectations adaptively interact during action prediction. There is also very little evidence concerning the brain mechanisms that underlie these adaptive interactions. The broader contribution of this work is to highlight the importance of behavioral constraints and simple decision heuristics in the emergence of cultural traditions.

In the **Experimental Contribution Chapter 1** entitled '**Perturbing the Action Observation Network during perception and categorization of others' actions: state-dependency and virtual lesion TMS effects**', we tested the respective contributions of the inferior frontal cortex (IFC), the anterior intraparietal region (AIP) (two areas that are thought to form part of the Action Observation Network (AON), also referred to as the human mirror system) and the primary somatosensory cortex (S1) in the perception and the recognition of observed tool-directed actions. To do so, we used a transcranial magnetic stimulation adaptation paradigm (TMSA). The aim of this type of paradigm is to manipulate, prior to the stimulation, the initial state of brain regions thought to play a role in the decoding a specific stimulus. The basic assumption is that if the target regions are involved in the decoding of test stimuli, the repeated exposure to a constant stimulus should result in an habituation in a subset of neurons that are located in these regions (Silvanto et al., 2008; Cattaneo et al., 2008). As neurons encoding the adapted attribute of the stimuli (i.e., the type of grip used versus the state of the effector configuration when the action outcome is reached) are made less active/excitable by adaptation, the application of TMS over the target regions should perceptually/behaviorally facilitate the less active/excitable neural populations relatively more than the active ones (e.g. faster reaction times for recognizing the stimulus that was previously adapted). After the adaptation phase, neurons that are sensitive to the stimulus attributes are assumed to have a baseline level of activity that is lower than neurons that are not tuned by adaptation (Li et al., 1993), and the latter group are more prone to reach a ceiling level of activation. This facilitation effect may occur because for the less active neurons there is a greater range for firing rate to be increased. TMSA paradigms enable one to estimate the causal involvement of brain regions in behavioral tasks (e.g. perceptual discrimination tasks) with better spatial resoluation than repetitive TMS techniques (see Silvanto and Pascual-Leone, 2008).

Using TMSA, we were interested in testing the causal involvement of the sensorimotor nodes of the AON in the differential visual coding of specific actions and action components. Participants were presented with adapting movies of an actor performing complex goal-directed actions on a tool (actions in which an object was lifted in order to open a box versus actions in which the same object was turned in order to switch-on a light) by using two kinds of grips (actions achieved with the use of a power versus precision grip) and were further asked to categorize test pictures as showing similar or different action/grips relative to the adapting movie. TMS was applied after the adaptation phase, at the onset of each test picture.

The key finding of this study was that applying TMS over S1 and IFC induced state-dependent effects on action recognition (see chapter 1 of the thesis, figure 4, pp. 98). TMS over S1 induced a selective decrease in the Inverse Efficiency index (i.e., a single measure of performance merging both RTs and accuracy) for pictures presenting the adapted action, indicating that stimulation of S1 improved the visual analysis of actions to which participants have been previously adapted. A similar improvement in performance for adapted relative to non-adapted actions was found with TMS over IFC, but not with sham stimulation nor stimulation of AIP. These results suggest that TMS over S1 and IFC specifically enhanced performance of the neural subpopulations that respond to a specific invariant feature, i.e. the type of arm action, between the adapting stimulus and the test stimulus. The TMS-induced behavioral enhancement occurred when subjects had to attend to such invariant feature (i.e. in the Arm Action recognition task) and was absent when processing of the same feature was task-irrelevant (i.e. in the Grip recognition task).

These findings, to the best of my knowledge, provide the first causative evidence of an involvement of S1 in the perception of complex goal-directed actions. This raises the question about whether S1, like AIP and IFC, contains mirror neurons that discharge both when an individual executes an action and when she/he observes the very same action (di Pellegrino et al., 1992). Interestingly, it has been demonstrated that half of the neurons of the IFC (specifically the ventral part of the premotor cortex) respond to somatosensory stimulation, suggesting that the mirror system may have functional links with the somatosensory cortices (Rizzolatti et al., 1988). S1 and S2 are known to be recruited in the processing of tactile, proprioceptive, and nociceptive information (see for a review Keysers et al., 2010). S1 has been shown to be more active when viewing hands manipulating objects (e.g., grasping a cup of tea) than actions that do not involve object manipulation (e.g., pointing movements) (Buccino et al., 2001; Pierno et al., 2009). Additionally, a differential activation of S1 was revealed when observers watched someone moving a heavy object compared to a light object (Molnar-Szakacs et al., 2006). Finally, it has been shown that when participants watched movies in which demonstrators manipulate objects, activity was consistently observed in S1. According to Keysers and co-workers (2010), this suggests that S1 could be involved in 'representing the haptic combination of tactile and proprioceptive signals that would arise if the participants manipulated the object in the observed way' (pp. 423).

The results of our study could be in line with these findings. However, the improvement in behavioral performance during the stimulation of S1 was observed when participants had to categorize the type of action that was performed by the filmed demonstrator, independently of the type of grip used to achieve these actions. One interpretation of this is that S1 actively participates in a mid-level analysis of the observed motor actions, taking into account the general state of the effector at the end-state of the action rather than the local variation of movement kinematics such as those characterizing the difference between power and precision grips. Possibly, S1 may inform other regions of the AON (and in particular the IFC in which action selection processes occur) with somatic cues that could be further used to 'simulate' the type of action that is currently observed. Unfortunately, our study failed to reveal consistent behavioral differences between the types of adapted actions. Indeed, according to the findings reported above, one may expect that stimulating S1 would improve the recognition of the adapted lifting action compared with the adapted switching action, especially when these two actions were performed with a grip that increased their general biomechanical costs (i.e., the use of a precision grip is better suited than a power grip to achieve the light switching action and, conversely, the use of a power grip is better suited than a precision to achieve the box opening action). Although S1 has been shown to be sensitive to the sensorimotor constraints associated with observed actions (Molnar-Szakacs et al., 2006), we did not find such an effect. The absence of such an effect in our experiment is not contradictory in itself, as it could suggest that S1 underlies our general capacity to represent what it would feel like to move one's own arm and hand in an observed way, independently of whether what is observed is associated with a high or low sensorimotor constraints. In conclusion, the observation of other people's actions recruits not only the classical sensorimotor nodes of the AON such as IFC and AIP, but also S1 which could be involved in 'simulating' how our own body would move and interact with objects that are the targets of the actions we observe.

The first experimental study of this thesis showed that action recognition requires the contribution of the primary somatosensory cortex, which is involved in the decoding of sensorimotor as well as somatic information conveyed by observed actions. Together, this information can be used by an observer to estimate the sensorimotor constraints (the muscular and articulatory costs) engaged in the execution of an action that is currently observed. In the **Experimental Contribution Chapter 2** entitled '**Object Affordances Tune Observers' Prior Expectations About Tool-Use Behaviors**', we postulated that the detection of visual cues – the affordances of an object or a tool – that enable an observer to make predictions about such constraints activate prior information that biases the interpretation of other people's actions. Such priors (here termed 'biomechanical priors') were thought to provide an economic, *by default*, interpretative strategy on which observers rely in order to understand and predict actions. Besides prior information conveyed by an estimation of the sensorimotor constraints of an observed actions, human observers can take advantage of another kind of prior information to infer other people's intentions, that is, information extracted from the statistical regularities of past events (Griffiths et al., 2008). It has been shown recently that human observers use the probability of occurence of another agent's intentions as a reliable source of information to infer, from observation, the intentions of an agent's upcoming actions (Chambon et al., 2011a, 2011b). Furthermore, these studies showed that participants were able to modulate the contribution of this information as a function of the available perceptual evidence as well as the type of intention to be inferred. More specifically, human observers progressively disengage from sensorimotor information in favor of probabilistic information when the perceptual evidence decreases (i.e., when the visual information conveyed by the action scene was stopped before the completion of the observed action) and when the goal to be inferred pertains to the higher levels of description of the action hierarchy (i.e., when the goal could potentially be achieved by a variety of motor actions).

The objective of our second experiment was thus to manipulate the biomechanical and probabilistic prior expectations of participants in order to study whether, and how, they interacted during action prediction. We designed a task in which participants were required to infer, under various conditions of visual uncertainty, the intentions of a demonstrator who manipulated a two-purpose tool using either a power or a precision grip in order to achieve two different intentions (opening the box versus switching-on the light). Affordance-related priors (termed 'biomechanical' priors) and priors acquired from past observations (termed 'probabilistic' priors) were manipulated by varying the biomechanical *optimality* (i.e., using a precision grip to switch-on the light and using the power grip to open the box minimized the sensorimotor constraints associated with goal achievement) of the tool-use actions and the *probability* (low versus high) of observing optimal versus suboptimal tool-use actions. Results of this study showed that biomechanical priors modulate the extent to which participants' predictions are influenced by probabilistically-induced expectations. Crucially, it was revealed that when the demonstrator's behavior satisfied both the participants' biomechanical and probabilistic priors, participants were able to efficiently combine both types of priors to make their predictions. Conversely, when the demonstrator's behavior conflicted with the biomechanical but not the probabilistic priors, it was costly for participants to inhibit the irrelevant sources of prior information, that is, the biomechanical priors. Nonetheless, our results suggest that increasing the number of observations of suboptimal behaviors could lead to a progressive decrease in the influence played by biomechanical priors, in favour of probabilistic priors.

Overall, our study demonstrates that prior information elicited by the sensorimotor constraints of the observed action and priors acquired from probabilistic exposure both contribute to action understanding. However, it appears that their respective weight in the prediction depends on the type of behavioral 'environment' (the type of behavior that is more likely to occur given past observation) observers are confronted with: biomechanical priors being particularly suited to make predictions in familiar behavioral environment (e.g., when observed agents behave according to the rules of biomechanical optimality); probabilistic priors being particularly suited to generate predictions in new or unfamiliar behavioral environment (e.g., when rules of biomechanical optimality are overridden or patently violated).

One of the intriguing aspects of this study concerns the strong predictive value of object affordances. Indeed, we propose that object affordances trigger a simple decision heuristic that is particularly efficient in predicting tool-use behaviors in conditions of visual uncertainty, and we observed that this decision heuristic is particularly difficult to inhibit. We proposed that affordances are predictive cues because they are perceived within the principle of rationality (Dennett, 1987), i.e., the fact that an action goal is expected to be achieved with the most optimal action means that are available given the situational constraints in which the action takes place (Gergely and Csibra, 2003; Csibra and Gergely, 2007). Here, object affordances are predictive of the upcoming action goals because each of them (the affordance that elicits a power grip and the affordance that elicits a precision grip) is 'rationally' adapted to reach a specific goal (e.g., the affordance that elicits a power grip is particularly prone to be exploited in order to open the box), with the rationality of the observed actions depending on whether or not the agent minimizes the muscular and articulatory costs.

The mere visual detection of an object affordance has been shown to automatically trigger in the observer's motor system (in the anterior intraparietal region and in the inferior frontal cortex of monkeys and humans) a set of motor commands corresponding to the afforded action (Murata et al., 1997; Murata et al., 2000; Gallese et al., 1994; Fogassi et al., 2001; Binkofski et al., 1998, 1999; Grèzes et al., 2003). Following this, one can rightfully question whether the predictive value of affordances evidenced by our results can be accounted for by a mechanism of motor 'simulation' or motor 'resonance' generated by the human mirror system. The answer to this question depends upon the function one attributes to the mirror system.

The first alternative – the well known *direct matching* hypothesis (Rizzolatti et al., 2001) –supports the idea that observing a demonstrator who is about to exploit a particular affordance activates in the observer's mirror system low-level motor representations corresponding to the detailed kinematics that are currently observed (i.e., the hand and digit configuration prior to the grasping movement). The activation of low-level motor representations by observation, resulting from visuomotor transformation carried out by mirror neurons, then propagates upwards in the observer's own hierarchically organized action system (see Hamilton and Grafton, 2006) to estimate which higher level goals might have generated the observed action (Fogassi et al., 2005; Iacoboni et al., 2005; Wolpert et al., 2003). However, this 'bottom-up' propagation would be efficient only for predicting the action's motor sub-goals (e.g., predicting the type grasping movement that is about to be performed), and not the higher level goals (e.g. opening the box) (Jacob and Jeannerod, 2005).

In our experiment, action stimuli were designed such that predicting the underlying intention poses an inverse problem (Baker et al., 2009), i.e., the analysis of sensorimotor information conveyed by the initial stages of the observed actions is not sufficient to unambiguously infer the demonstrator's intention. Thus, how do observers, when confronted with a specific type of grasping movement, select a particular goal among the two possible alternatives (i.e., opening the box versus switching-on the light)? Unfortunately, the *direct-matching* hypothesis alone fails to explain the directionality of the participants' goal choice, whether it is biased by probability or by the observed biomechanical constraints. Nonetheless, it does not mean that mirror mechanisms are not involved in this kind of inference.

Mirror mechanisms may indeed be recruited, but in a different way. For the motor simulation mechanisms to be involved in the prediction of complex goal-directed actions such as those presented in our second experimental work, one should assume that the observer has a prior representation of the goal that is more likely to be achieved by the demonstrator (Csibra, 2007; Kilner et al. 2007a, 2007b; Kilner, 2011). This is precisely what the rationality principle presupposes: considering an observed action as rational (or, in the context of our task, as biomechanically optimal) means that a causal link is drawn between the action means (e.g., using a precision grip) and the goals (e.g., turning the tool in order to switch-on the light). Thus, watching a demonstrator using a specific object affordance enables an observer to select *by default,* between the two concurrent goals, the one that minimizes the muscular and articular costs. This selection can occur because the observer has a prior representation of these two concurrent goals. This strategy has been showed to be particularly efficient when participants have no additional information about the general behavior of the demonstrator (i.e., information acquired from a probabilistic sampling of past events), or when the visual information conveyed by the action scene is noisy or incomplete. We showed that when participants had such additional information (i.e., when they integrated the probabilistic bias in which suboptimal behaviors were favoured), they were able to progressively inhibit the affordance-based inferential strategy to favor a strategy based on probabilistic information, though to a lesser extent. In this context, action mirroring would serve a very different function from the function postulated by the *direct-matching* hypothesis. This function is not to access the high-level goals and intentions through bottom-up propagation but to anticipate the course of the observed action through a top-down *reconstruction* of its motor sequence (Csibra, 2007; Kilner et al., 2007a, 2007b, Kilner, 2011). Crucially, this means that the understanding of the action goal and intention is the *input* rather than the *output* of the mirroring process. According to the *'action reconstruction'* hypothesis, simulation occurs as the prior representation of the overarching goal is mapped onto the observer's own mirror system, within which it can propagate downwards to generate the corresponding motor code at the lower levels (see also the chapter E of the General Introduction of the thesis). According to Csibra (2007), this top-down propagation does not contradict the idea of motor simulation but instead, assumes that motor simulation is 'predictive in nature, generating motor actions for goal conjectures rather than the other way around' (pp. 441).

In summary, these two models of action mirroring can be differentiated by two aspects: the action interpretation level at which visuomotor transformation is performed (low-level versus high-level mirroring) and the directionality of the propagation within the action system following mirroring (bottom-up versus top-down propagation). The *'direct-matching'* hypothesis assumes that simulation mechanisms occur at a low-level of action description (i.e., the kinematic level) and generate a bottom-up propagation within the hierarchically organized action representation system, supposedly allowing the inference of high-level goals and intentions. The *'action reconstruction'* hypothesis assumes that simulation mechanisms occur at a high or intermediate level of description (i.e., the intention or the goal level) and generate a top-down propagation within the hierarchically organized action interpretation system. This top-down propagation enables the on-line monitoring of the action course in order to test the likelihood of the *to-be-predicted* intentions. It is unlikely that the first interpretation fits the results obtained in our second experimental work, especially because *direct-matching* fails to solve the inverse problem represented by the understanding of our action stimuli (one type of grasping movement could equally lead to two action goals). The second interpretation, however, could fit with these results, especially because participants based their prediction on a prior representation of the demonstrator's intentions, being inferred from a rational estimation of the biomechanical costs (low versus high) of the observed actions (biomechanical priors) or from the probabilistic exposure (probabilistic priors) to a particular behavior (biomechanically optimal or suboptimal).

The *‘action reconstruction’* hypothesis – for which the predictive coding model (see chapter E of the General Introduction) developped by Kilner and co-workers provides a biologically plausible framework (Kilner et al., 2007a, 2007b; Kilner, 2011) – also offers an interesting framework to interpret some results obtained in the third experimental work presented in this thesis.

The first objective of the **Experimental Contribution Chapter 3**, entitled ‘**Modulating human motor resonance: exposure to suboptimal actions suppresses corticospinal excitability’,** was to exploit the behavioral results obtained in the experimental chapter 2 and to assess whether the motor system could hold traces of the behavioral changes exhibited by the observed demonstrator. In order words, I investigated whether the interactions between biomechanical priors (conveyed by the detection of object affordances) and prior knowledge (acquired from probabilistic exposure) about a demonstrator’s behaviors, which influence the participants’ predictions of the demonstrator’s intentions (see Experimental Contribution Chapter 2) can modulate the motor system activity. To this aim, I used single-pulse TMS applied over the primary motor cortex (M1) to measure of the corticospinal excitability (CSE) of participants during action prediction. To do so, we adapted the experimental design used in our previous study (see Experimental Contribution Chapter 2) to an on-line TMS paradigm. Once again, the biomechanical optimality of tool behaviors performed by the demonstrator, as well as the probability of observing him achieving an intention using optimal and suboptimal behaviors, were varied. Three behavioral ‘environments’ were then created, each characterized by a different probability of observing the demonstrator performing optimal and suboptimal behaviors (i.e., in the no bias condition, the demonstrator equally performed optimal and suboptimal behaviors while in the convergent bias session the demonstrator favoured optimal behaviors and in the divergent bias session he favoured suboptimal behaviors). While the convergent probabilistic bias provided a stable environment (i.e., an environment that conformed to the predictions based on biomechanical priors), the no bias and the divergent probabilistic biases provided comparatively more unexpected, open-ended environments (i.e., an environment that did not conform to the predictions based on biomechanical expectations). We applied TMS over the left M1 of participants during the prediction of visually uncertain actions, both before and after the probabilistic exposure.

Two lines of results emerged. Behaviourally, we closely replicated our previous findings by showing that when the behaviors gathered from probabilistic exposure and biomechanical prior expectations of participants converged, they efficiently combined both types of priors to make their predictions. Conversely, when the two priors diverged, participants had to deal with two conflicting sources of prior information, which resulted in predictions close to chance level. Physiologically, we showed that the the type of probabilistically-induced behavioral environment translated into changes in CSE. Specifically, a decrease in CSE occurred when the participants were exposed to behaviors that violated the biomechanical optimality rules, i.e., when the observed action could not be matched with the participant’s biomechanical expectations. These last results reveal that CSE, which possibly reflects the involvement of motor resonance mechanisms in action observation (Fadiga et al., 2005) is malleable and can be altered IN TIME, by varying the degree to which biomechanical and probabilistic prior expectations match. Indeed, high levels of CSE were maintained only in environments that were consistent with the convergence of biomechanical and probabilistic prior expectations (in our task: the convergent bias session).

Interestingly, these results suggest that the acquisition of probabilistic information affects the level of CSE, and potentially, the state of the mirror system. However, CSE changes were only observed at the group level, that is, according to the type of probabilistic bias participants were exposed to. Furthermore, no differences in CSE were revealed between the different types of action that participants observed, both during the first and the second series of visually uncertain actions. The absence of a difference is not surprising in itself. One can indeed consider that motor simulation mechanisms are not a *direct-matching* but instead, an *action reconstruction* process that starts from an analysis of the higher-level action component that is then propagated downwards through the hierachically organized action representation system to generate the lower level motor codes (Csibra, 2007; Kilner et al., 2007a, 2007b; Kilner, 2011). In the three groups of participants, the CSE level measured before the probabilistic exposure was increased relative to the observation of a black screen. CSE remained high after exposure to the no bias and convergent bias, but exposure to the divergent bias significantly decreased CSE levels. This might be because during the incomplete movies that preceded the exposition to the probabilistic biases as well as during the incomplete movies that followed the convergent bias, participants' predictions were mainly driven by their biomechanical priors. In other words, they expected to observe an action sequence that was congruent with their prior representation of the goals, that is, an action sequence that minimized the biomechanical costs given the final goal that had to be reached. More specifically, the level of CSE facilitation was relatively constant within these conditions possibly because in each of them, the likelihood of observing an intention resulting in optimal behaviors was confirmed both by the observation of complete movies and by their biomechanical priors – priors that were used *by default* during the incomplete movies to complement the missing visual information (Jacquet et al., 2012b, in press). However, CSE decreased after exposure to suboptimal behaviors possibly because the likelihood of observing optimal behaviors progressively decreased as participants accumulated disconfirmatory evidence. The decrease in CSE could potentially occur because of the mismatch between what was actually observed by participants (a demonstrator achieving his intentions using suboptimal behaviors) and their prior expectations derived from biomechanical constraints of the motor system.

Interstingly, such an interpretation is consistent with the 'predictive coding' framework proposed by Kilner and co-workers (Kilner et al., 2007a, 2007b, 2011) which proposes a biologically plausible model for the action reconstruction hypothesis. The 'predictive coding' framework postulates that hierarchically distinct cerebral regions (having 'mirror' properties or not, such regions forming the AON network) subserve the different levels of action understanding: the *kinematic* level, the *motor* level, the *goal* level, and the *intention* level. The model postulates that each level of the hierarchy at which an action can be analysed generates prediction signals that specify, or bias, the representations at the inferior level. Predictions generated at the highest levels are then compared with current predictions generated at the lower levels. Thus, depending on the type of goal being anticipated, the observer will predict a motor command that is congruent with this goal and, on the basis of her own motor repertoire, will generate expectations about the specific kinematics that best fit with the predicted motor command. The comparison between the predicted and the currently observed kinematics will generate, in the case of a mismatch, a prediction error. The magnitude of such a prediction error will then be reduced, via *forward* connections, by updating the representation elaborated at the more abstract, higher levels. This exchange of reciprocal signals proceeds until the cause that most likely explains the observed action is inferred, or, in other words, until the magnitude of the prediction error is sufficiently minimized to enable accurate inference (Kilner et al., 2007a; 2007b). According to this model, the mirror neuron system would be particularly active when the prediction error is minimized. Interestingly, the level of CSE in our study decreased after expectations generated by*observation* and prior *expectations* conflicted (divergent bias), and this could possibly generate a prediction error resulting in the weakening of the mirror activity.

This influence of prior expectations could be conceived as a way to either facilitate or inhibit mirror system activity, depending on whether the probabilistically-induced ‘behavioral’ environment converges or diverges with the sensorimotor constraints of the system. Our findings support the second possibility (i.e., lack of resonance when the behavior biased by probabilistic exposure diverges from sensorimotor constraints). Together, these findings are crucial as they suggest that mirror processes are not automatically engaged during action understanding or, to say the least, that these processes are *altered* by higher-order expectations about other people’s intentions and behaviors.

One may speculate about whether this alteration of CSE is purely incidental and passive, resulting from a mismatch between what is observed and the motor programs that are stored in the observer’s own motor reportoire, or whether it reflects an active inhibitory process. Being equipped with active inhibitory process would provide certain advantages. The suppression of motor resonance through higher-order probabilistic representations of the environment could be viewed as a subtle adaptive response, for in some environments mirror mechanisms would lose their predictive value. Thus, our argument is that such modulatory effects may protect the observer against maladaptive predictions in the case where observation and expectations conflict. Further studies are needed to establish whether CSE suppression is obtained via a passive process (i.e., conditions violated = no resonance) or implies an active inhibitory process that could veto ‘automatic’ motor resonance activity in order to reduce the weight of inadequate action representations (Khün et al., 2009).

Ultimately, the active regulation of motor resonance processes through higher-order probabilistic representations of the environment may provide an adaptive mechanism to enable acquisition of unexpected, new behaviors. Interestingly, behaviors that override rules of biomechanical optimization are regularly promoted by human culture. Relying on prior knowledge gathered from probabilistic sampling of past observations may thus occasionally prove more helpful than merely evaluating the (biomechanical) optimality of the observed behavior.

The first three experimental contributions presented in this thesis were previously discussed highlight the role of sensorimotor (including somatic) constraints and prior information in the ability of human observers to understand their conspecifics' intentions. I attempted to illustrate how these types of information may generate distinctive decisional biases that alter the prediction of observed actions. In particular, the extraction of information concerning the frequency of occurrence of a particular intention was efficiently used by observers to make inferences about intentions underlying up-coming actions. Additionally, the detection of sensorimotor constraints from visual cues such as object affordances activates a simple decision heuristic that, coupled with expectations about the rationality of observed actions, provides a *by default* strategy upon which observers rely to make predictions about intentions of observed agents. Our experimental works show that these two decisional mechanisms could be differentially recruited according to the type of 'behavioral' environments observers are confronted with: the probabilistic inference being particularly adapted to make predictions in unfamiliar 'behavioral' environments (i.e. when the likely behaviors are biomechanically suboptimal), and the affordance-based heuristic being particularly adapted to make predictions in familiar 'behavioral' environments (i.e., when the likely behaviors are biomechanically optimal). Nevertheless, I believe that one of the most interesting findings of these works is the strong predictive value of object affordances, especially because it represents a cognitively tractable and economic inferential strategy that is efficient in a numerous daily situations. It provides an alternative to complex, cognitively costly strategies that are usually evoked by researchers to account for the social learning of complex object-directed actions such as tool use.

I propose that affordances, and the function of minimizing biomechanical costs, have a significant impact on the acquisition of behaviors through social learning. In a recent commentary article (Jacquet et al., 2012a, in press), I addressed this question to Krist Vaesen (Vaesen, 2012) who, in his target article, argued that the acquisition of tool-use behaviors from social learning crucially depended on uniquely human, sophisticated socio-cognitive skills. My purpose was to claim that those skills were also based on simpler detection systems humans could share with other animal tool users. Accordingly, I discussed the impact of object affordances on the understanding and the social learning of tool use.

Here I report the main gist of this commentary (see also appendix B for the complete version) which, I believe, provides a clear example of how the results of the experimental chapter 2 and 3 can be extended to the broader domains of social learning and cultural evolution:

> "Krist Vaesen speculates that the humans' capacity to learn novel tool use from observing goal-directed movements performed by others (Csibra & Gergely 2007) is a hallmark of our uniqueness, and is based on "higher" socio-cognitive skills. It has been proposed that such skills were supported by the ability to i) decode kinematic information into causal relationships between a behavioural sequence and its result (Gergely 2007), ii) interpret others' behaviors as rational (assuming that the most efficient observed action means are adopted to achieve a particular goal; Gergely & Csibra 2003) and iii) accumulate *a priori* knowledge from past observations about

agents‟ intentions and behaviours in order to predict future events (Chambon et al. 2011). We agree with the author that the sophistication of such socio-cognitive skills goes far beyond those of any other animals. Yet, we believe that this sophistication could also be the result of simpler systems allocated to the detection of low-level, local sources of information, such as the manipulative properties of objects called “affordances”.

Affordances define relational properties that emerge from matching the perceived physical features of objects and the agent’s biomechanical architecture, goals, plans, values, beliefs, and past experiences. We propose that affordances allow agents to delineate the number of candidate motor acts that could be performed on tools. We postulate that affordances constrain the number of possible solutions by generating biomechanical prior expectations in line with the bodily architecture of agents. These priors would bias individuals to act towards objects aiming at biomechanical optimization (Rosenbaum et al. 1996; Weiss et al. 2007). As the author rightly points out, compared to other animals, the many degrees of freedom characterizing human effectors and their striking motor control considerably enhances our ability to detect new affordances and new potential objects uses. All this contributes to increase the variety of the behavioural repertoire.

Nonetheless, we are skeptical about the idea that the primary advantages such architectural properties brings for tool use acquisition is fine-grained social learning. Indeed, in many situations, detecting tools affordances allows learners to avoid such a high-level but costly strategy. Instead, this biomechanical uniqueness could increase the probability of individual innovation, particularly in situations where novel tools are physically unstructured and multi-purpose. For example, Acheulean stone tools are poorly structured and roughly symmetrical objects with a cutting edge. They do not offer affordances salient enough to constrain the number of candidate motor acts that could be perform on them. Sterelny (2003a) points out that the exact functions and uses of Acheulean stone tools, though they were the dominant element of human technology for over a million years, remain a matter of debate. It is more plausible that our ancestors – who were predisposed to behavioural innovation thanks to their high biomechanical flexibility – progressively discovered not one or two, but a multitude of tasks that Acheulean stone tools could roughly carry out.

We argue that the evolution of the human technological environment favoured the utility of simpler systems such as affordances detection. This eases the negotiation of the highly demanding cognitive problems of tool use learning (Clark 1997; Dennett 1995; Sterelny 2003a, 2003b). Indeed, tools we interact with daily are designed for specific purposes. Affordances that are available through their complex physical attributes offer the chance for naive users to extract their functions at low cost (Dennett 1982, 1995; Gregory 1981; Norman 1988). In our engineered environments, affordances play a crucial role in the acquisition of tool skills through individual trial-and-error as well as social learning. More specifically, we argue that perceiving affordances directly biases the understanding of tool behaviours performed by others, and consequently the extraction of related functional knowledge. The biomechanical priors that emerge from the perception of tools affordances constrain the number of candidate motor acts an individual could initiate. Similarly, they also tune the observer's prior expectations about which motor behaviors are most likely to be performed by others, enhancing their predictability and learnability. Learning about a novel tool from observing a demonstrator using it in a biomechanically "rational" way would be less costly than learning from a demonstrator that violates our expectations. That is, the convergence of the demonstrator's and observer's biomechanical expectations facilitates an efficient learning strategy, based on kinematics, rationality principle or prior knowledge. Taken together, these observations question the exact role of high-level, fine-grained social learning in the acquisition of new tool skills. Relevant to this is work addressing animal behavioural "traditions" – behavioural patterns which are relatively stable in groups and are at least partly maintained by some forms of social learning. These could result from constraints that limit the number of possible alternative behaviours, more than from the robustness of high-level social transmission mechanisms (Claidière & Sperber 2010; Tennie et al. 2008). Here, we posit that the crucial role affordances play in the acquisition of tool use strongly suggests that fine-grained social learning strategies, such as true imitation of observed action goals and means, is sometimes less important than previously assumed. In fact, affordances, together with ecological constraints and other products of epistemic engineering, could enhance the effectiveness of more frugal forms of socially-directed learning (Acerbi et al. 2011; Franz & Matthews 2010) like emulation learning (i.e. the observer copies action goals performed by a demonstrator without considering action means) or even stimulus enhancement (i.e. when an individual directs its behaviour towards an object or a part of an object with which it saw another individual interact).”

In the above commentary, I speculated that sensorimotor constraints could bias the learning of new behaviors from observation. The **Experimental Contribution Chapter 4** entitled ‘**Behavioral constraints and the evolution of faithful social learning**’ examined, with the help of an individual-based model, the relationship between similar constraints and the evolution of faithful social learning. The claim of this work is that the emergence of behavioral traditions within a population of human and non-human individuals could be accounted for by both high-fidelity (e.g., imitation and emulation) and low-fidelity (e.g., stimulus enhancement) social transmission mechanisms. The model aimed to demonstrate that the use of high-fidelity and low-fidelity social learning strategies was a function of both the type of behavioral constraints (biomechanical and cognitive factors that specify the number of potential behaviors individuals were provided with) and the type of search space of a task (the number of alternatives observers could perform to reproduce the outcome of the behavior performed by the demonstrator). We thus tested this hypothesis by modeling three fictive populations of individuals, each characterized by a different number of available behaviors (10 versus 100 versus 200 possible alternative behaviors). Each population was then submitted to three specific task problems, each being potentially solved by a certain number of behavioral alternatives. More specifically, in the *smooth space*, many behavioral alternatives could be achieved to solve the task problem. In the *rugged space* however, only few alternatives could lead to the optimal solution. Finally, in the *peaked space*, just one alternative led to the optimal solution. These three task structures were thought to differentially orient the individual search (trial and error learning), in such a way that for the *rugged* and *peaked* space, social-learning was expected to be more efficient than individual learning (for only few solutions led to payoffs). Thus, we simulated interactions at the level of single individuals, such that individuals of a population interacted in discrete time-steps. At each time-steps individuals interacted in pairs. Each individual (the *observer)* was paired with another individual (the *demonstrator*) randomly chosen among the ten individuals of the population with the highest payoffs (the payoff an individual received from performing a behavior). Thus, each learner attempted to copy the behavior of its paired demonstrators. An evolutionary algorithm was further used to optimize the behavior of individuals. The evolutionary algorithm acted on a variable that encoded the fidelity of social learning of each individual. Individuals that performed better (i.e., that maximize the fidelity of the copy) resulted in proportionally more “offspring” than others. Finally, we also added the possibility of environmental changes by assigning different probability for the three distributions payoff (smooth versus rugged versus peaked space) to be reinitialized during the simulations. This procedure allowed analyzing how the interaction between the fidelity of social learning and the size of the behavioral repertoire impacted the populations’ behavioral diversity, namely, the number of behavioral patterns present in a population. Three main results emerged from the simulations. First, in stable environments (environments where the payoff distribution remained unchanged), faithful social learning evolved only in populations with large behavioral repertoires, and particularly in peaked search spaces. Second, results showed that the convergence towards behavioral homogeneity resulted from high-fidelity social learning mechanisms but also from low-fidelity mechanisms. Third, intermediate rates of environmental change favored the evolution of faithful social learning.

These results are particularly relevant for the take-home message of the present thesis. Indeed, our simulations suggest that a population of individuals that is highly constrained (small number of potential behaviors) can evolve something similar to cultural traditions, without the need of costly faithful social-learning abilities such as true imitation. The fact that this could happenin peaked space (i.e., a search space that is unlikely to favor individual learning, given the small number of solutions that bring a payoff) is important. Indeed, it could mean that, even with human populations, many complex behaviors can be 'transmitted' from an individual to another without the need of sophisticated social-learning abilities. This is the idea I speculated on in the commentary article previously mentioned (section B of the general discussion or appendix B). I believe that this is precisely what happens when a human demonstrator learns object-directed or tool-use actions by observing a third party. The observer's behavior is guided by the interaction between the object affordances and the biomechanical constraints of his body, such that both the observer and the demonstrator's behavior converge. Consequently, the outcome of the learning process resembles a sophisticated imitative process but, instead, reflects a convergence that results from the processing of sensorimotor cues. This convergence would be incidental, in the sense that it does not result from a sophisticated social learning mechanisms.

Finally, I argue that similar phenomena occur more frequently than usually assumed in our human technological societies. Indeed, even very complex human 'cultural' behaviors such as tool-use and other object manipulations may be a result of a combination of a constrained behavioral repertoire, constrained search spaces, and some form of social learning (Sperber, 1996). This is possible because humans engineer their environment in such a way that it becomes informationally transparent (Sterelny, 2003). For example, humans deliberately manufacture tools whose complex physical attributes offer naïve users informative cues (the affordances) that enable the extraction of their functions at low cost (Dennett, 1982, 1995; Gregory, 1981). I believe that the evolution of human technology might have favored the use of low-fidelity social transmission mechanisms, and so, without affecting the final fidelity of the transmission.

REFERENCES

Acerbi, A., Tennie, C., & Nunn, C. L. (2011). Modeling imitation and emulation in constrained search spaces. *Learning & behavior*, *39*(2), 104-114. doi:10.3758/s13420-010-0009-z

Acerbi, A., Jacquet, P.O., Tennie, C. (2012). Behavioral constraints and the evolution of faithful social learning. *Current Zoology, 58*(02), 307-318.

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature neuroscience*, *11*(9), 1109-1116.

Akins, C. K., & Zentall, T. R. (1996). Imitative learning in male Japanese quail (Coturnix japonica) using the two-action method. *Journal of comparative psychology (Washington, D.C.: 1983)*, *110*(3), 316-320.

Alaerts, K., Senot, P., Swinnen, S. P., Craighero, L., Wenderoth, N., & Fadiga, L. (2010). Force requirements of observed object lifting are encoded by the observer's motor system: a TMS study. *The European journal of neuroscience*, *31*(6), 1144-1153. doi:10.1111/j.1460-9568.2010.07124.x

Alaerts, K., Swinnen, S. P., & Wenderoth, N. (2009). Is the human primary motor cortex activated by muscular or direction-dependent features of observed movements? *Cortex; a journal devoted to the study of the nervous system and behavior*, *45*(10), 1148-1155. doi:10.1016/j.cortex.2008.10.005

Allison, Puce, & McCarthy. (2000). Social perception from visual cues: role of the STS region. *Trends in cognitive sciences*, *4*(7), 267-278.

Asch, S.E. (1955). Opinions and social pressure. *Scientific American*, 193(5): 31–35.

Asch SE (1951) Effects of group pressure upon the modification and distortion of judgments. In: Guetzkow HS, ed. *Groups, Leadership and Men Research in Human Relations*. Pittsburgh: Carnegie Press, 177–190.

Aziz-Zadeh, L., Maeda, F., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2002). Lateralization in motor facilitation during action observation: a TMS study. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *144*(1), 127-131. doi:10.1007/s00221-002-1037-5

Baker, C.L., Tenenbaum, J.B., Saxe, R.R. (2006). Bayesian models of human action understanding. *Advances in Neural Information Processing Systems 18* (pp. 99-106).

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329-349. doi:10.1016/j.cognition.2009.07.005

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a « theory of mind »? *Cognition*, *21*(1), 37-46.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., & Perrett, D.I., 2005. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, 17, 377–391.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245-249. doi:10.1038/nature07538

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current biology: CB, 18*(10), 775-780. doi:10.1016/j.cub.2008.04.051

Binkofski, F., Buccino, G., Stephan, K. M., Rizzolatti, G., Seitz, R. J., & Freund, H. J. (1999). A parieto-premotor network for object manipulation: evidence from neuroimaging.

Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale, *128*(1-2), 210-213.

Binkofski, F., Dohle, C., Posse, S., Stephan, K. M., Hefter, H., Seitz, R. J., & Freund, H. J. (1998). Human anterior intraparietal area subserves prehension: a combined lesion and functional MRI activation study. *Neurology*, *50*(5), 1253-1259.

Biro, S., & Leslie, A. M. (2007). Infants' perception of goal-directed actions: development through cue-based bootstrapping. *Developmental science*, *10*(3), 379-398. doi:10.1111/j.1467-7687.2006.00544.x

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature reviews. Neuroscience*, *2*(8), 561-567. doi:10.1038/35086023

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Current biology: CB*, *17*(24), 2117-2121. doi:10.1016/j.cub.2007.11.057

Brighina, F., La Bua, V., Oliveri, M., Piazza, A., & Fierro, B. (2000). Magnetic stimulation study during observation of motor tasks. *Journal of the neurological sciences*, *174*(2), 122-126.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *The European Journal of Neuroscience*, *13(2)*, 400-404.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current biology: CB*, *20*(13), 1165-1170. doi:10.1016/j.cub.2010.04.055

Carruthers, P. (1996). Simulation and self-knowledge: a defence of the theory-theory. In P. Carruthers & P.K. Smith, Eds. *Theories of theories of mind.* Cambridge: Cambridge University Press.

Catmur, C., Walsh, V., & Heyes, C. (2007). Sensorimotor learning configures the human mirror system. *Current biology: CB*, *17*(17), 1527-1531. doi:10.1016/j.cub.2007.08.006

Cattaneo, L., Barchiesi, G., Tabarelli, D., Arfeller, C., Sato, M., & Glenberg, A. M. (2011). One's motor performance predictably modulates the understanding of others' actions through adaptation of premotor visuo-motor neurons. *Social cognitive and affective neuroscience*, *6*(3), 301-310. doi:10.1093/scan/nsq099

Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010b). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(9), 2252-2258. doi:10.1093/cercor/bhp291

Cattaneo, Z., Devlin, J. T., Vecchi, T., & Silvanto, J. (2009). Dissociable neural representations of grammatical gender in Broca's area investigated by the combination of satiation and TMS. *NeuroImage*, *47*(2), 700-704. doi:10.1016/j.neuroimage.2009.04.097

Cattaneo, Z., Rota, F., Walsh, V., Vecchi, T., & Silvanto, J. (2009b). TMS-adaptation reveals abstract letter selectivity in the left posterior parietal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *19*(10), 2321-2325. doi:10.1093/cercor/bhn249

Cesari, P., Pizzolato, F., & Fiorio, M. (2011). Grip-dependent cortico-spinal excitability during grasping imagination and execution. *Neuropsychologia*, *49*(7), 2121-2130. doi:10.1016/j.neuropsychologia.2011.04.019

Chambon, V. (2010). Modèles hiérarchiques des représentations d'actions: apports de le psychopathologie et de la Neuroimagerie par Résonance Magnétique Fonctionnelle. Editions Universitaires Européennes.

Chambon, V., Domenech, P., Pacherie, E., Koechlin, E., Baraduc, P., & Farrer, C. (2011a). What are they up to? The role of sensory evidence and prior knowledge in action understanding. *PloS one*, *6*(2), e17133. doi:10.1371/journal.pone.0017133

Chambon, V., Pacherie, E., Barbalat, G., Jacquet, P., Franck, N., Farrer, C. (2011b). Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia. Brain, 134(12), 3728-3741.

Chase, V. M., Hertwig, R., & Gigerenzer, G. (1998). Visions of rationality. *Trends in cognitive sciences*, *2*(6), 206-214.

Chersi, F., Ferrari, P.F., & Fogassi, L. (2011). Neuronal chains for actions in the parietal lobe: a computational model. *PLoS One, 6(11):e27652.*

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239. doi:10.1016/j.tics.2012.02.007

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1485), 1585-1599. doi:10.1098/rstb.2007.2054

Claidière, N., & Sperber, D. (2010). Imitation explains the propagation, not the stability of animal culture. *Proceedings. Biological sciences / The Royal Society*, *277*(1681), 651-659. doi:10.1098/rspb.2009.1615

Clark, A. (1997). Being there: putting brain, body, and world together again. Cambridge, MA: MIT Press.

Clark, A. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioural and Brain Sciences, in press.*

Clark, S., Tremblay, F., & Ste-Marie, D. (2004). Differential modulation of corticospinal excitability during observation, mental imagery and imitation of hand actions. *Neuropsychologia*, *42*(1), 105-112.

Cohen Kadosh, R., Muggleton, N., Silvanto, J., & Walsh, V. (2010b). Double dissociation of format-dependent and number-specific neurons in human parietal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *20*(9), 2166-2171. doi:10.1093/cercor/bhp273

Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press.

Csibra, G, Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of « pure reason » in infancy. *Cognition*, *72*(3), 237-267.

Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 447-458. doi:10.1098/rstb.2002.1235

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1567), 1149-1157. doi:10.1098/rstb.2010.0319

de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary systems for understanding action intentions. *Current biology: CB*, *18*(6), 454-457. doi:10.1016/j.cub.2008.02.057

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature neuroscience*, *8*(11), 1611-1618. doi:10.1038/nn1575

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992b). Understanding motor events: a neurophysiological study. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *91*(1), 176-180.

Dennett, D. (1971). Intentional systems. *Journal of Philosophy*, 68, 87–106.

Dennett, D.C. (1982). Beyond beliefs. In: Woodfield A, ed. *Thought and object: essays on intentionality*. pp 1-97. Oxford, UK: Oxford University Press.

Dennett, D. (1991). Real patterns. *Journal of Philosophy*, 87, 27–51.

Dennett, D.C. (1987). The intentional stance. MIT Press, Cambridge, MA.

Dennett, D. (1995). Darwin's Dangerous Idea. Evolution and the Meanings of Life. New York: Simon and Schuster.

Dorrance, B. R., & Zentall, T. R. (2001). Imitative learning in Japanese quail (Coturnix japonica) depends on the motivational state of the observer quail at the time of observation. *Journal of comparative psychology (Washington, D.C.: 1983)*, *115*(1), 62-67.

Ellis, R., & Tucker, M. (2000). Micro-affordance: the potentiation of components of action by seen objects. *British journal of psychology (London, England: 1953), 91 (Pt 4)*, 451-471.

Fadiga, L, Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology*, *73*(6), 2608-2611.

Fadiga, Luciano, Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current opinion in neurobiology*, *15*(2), 213-218. doi:10.1016/j.conb.2005.03.013

Faillenot, I., Toni, I., Decety, J., Grégoire, M. C., & Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: functional anatomy with PET. *Cerebral cortex (New York, N.Y.: 1991)*, *7*(1), 77-85.

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of experimental psychology. Learning, memory, and cognition*, *28*(3), 458-467.

Fiser, J., & Aslin, R.N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15822-15826.

Flash, T., & Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. *Current opinion in neurobiology*, *15*(6), 660-666. doi:10.1016/j.conb.2005.10.011

Fogassi, L, Gallese, V., Buccino, G., Craighero, L., Fadiga, L., & Rizzolatti, G. (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: A reversible inactivation study. *Brain: a journal of neurology*, *124*(Pt 3), 571-586.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G.(2005). Parietal lobe: from action organization to intention understanding. *Science, 308(5722),* 662-667.

Franz, M., & Matthews, L. J. (2010). Social enhancement can create adaptive, arbitrary and maladaptive cultural traditions. *Proceedings. Biological sciences / The Royal Society*, *277*(1698), 3363-3372. doi:10.1098/rspb.2010.0705

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of physiology, Paris*, *100*(1-3), 70-87. doi:10.1016/j.jphysparis.2006.10.001

Frith, C D, & Frith, U. (1999). Interacting minds--a biological basis. *Science (New York, N.Y.)*, *286*(5445), 1692-1695.

Frith, Chris D, & Frith, U. (2006). How we predict what other people are going to do. *Brain research*, *1079*(1), 36-46. doi:10.1016/j.brainres.2005.12.126

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 459-473. doi:10.1098/rstb.2002.1218

Gallese, V, Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996b). Action recognition in the premotor cortex. *Brain: a journal of neurology*, *119 (Pt 2)*, 593-609.

Gallese, V, & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gallese, V, Murata, A., Kaseda, M., Niki, N., & Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. *Neuroreport*, *5*(12), 1525-1529.

Gallese, Vittorio. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 517-528. doi:10.1098/rstb.2002.1234

Gallese, Vittorio, Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in cognitive sciences*, *8*(9), 396-403. doi:10.1016/j.tics.2004.07.002

Gangitano, M., Mottaghy, F. M., & Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, *12*(7), 1489-1492.

Gergely, G, Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165-193.

Gergely, G, Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755. doi:10.1038/415755a

Gergely, G, & Csibra, G. (2003). Teleological reasoning in infancy: the nai;ve theory of rational action. *Trends in cognitive sciences*, *7*(7), 287-292.

Gibson, J.J. (1979). The ecological approach to visual perception. Boston, MA: Houghton Mifflin.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451-482. doi:10.1146/annurev-psych-120709-145346

Goldman, A.I. (1995). In defense of the Simulation Theory. In M. Davies et R. Stone. *Mental simulation: Philosophical and Psychological Essays*. Oxford : Blackwell.

Gopnik, A. (1993). How we know our minds : the illusion of first-person knowledge of intentionality. *Behavioral and Brain sciences*, 16, pp. 1-14.

Gopnik, A. & Meltzoff, A.N. (1994). Minds, bodies and persons: Young children's understanding of the self and others as reflected in imitation and 'theory of mind' research. In S. Parker et R. Mitchell, Eds. *Self-awareness in animals and humans*. New York: Cambridge University Press.

Gopnik, A., & Wellman, H.M. (2012). Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory. *Psychological Bulletin, in press*.

Gordon, R.M. (1996). Radical simulationism. In P. Carruthers et P. Smith, Eds. *Theories of Theories of Mind.* Cambridge : Cambridge University Press, pp. 11-21.

Grafton, S. T., & Hamilton, A. F. de C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human movement science*, *26*(4), 590-616. doi:10.1016/j.humov.2007.05.009

Gregory, R.L. (1981). Mind in science: a history of explanations in psychology and physics. Cambridge, UK: Cambridge University Press.

Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003a). Activations related to « mirror » and « canonical » neurones in the human brain: an fMRI study. *NeuroImage*, *18*(4), 928-937.

Grèzes, J., Armony, J. L., Rowe, J., & Passingham, R. E. (2003b). Activations related to « mirror » and « canonical » neurones in the human brain: an fMRI study. *NeuroImage*, *18*(4), 928-937.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition, In Ron Sun (ed.), *The Cambridge handbook of computational cognitive modelling*, Cambridge: Cambridge University Press.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in cognitive sciences*, *14*(8), 357-364. doi:10.1016/j.tics.2010.05.004

Grimm, V., Railsback. S.F. (2005). Individual-based Modeling and Ecology. Princeton, NJ: Princeton University Press.

Grossman, E. D., & Blake, R. (2002). Brain Areas Active during Visual Perception of Biological Motion. *Neuron*, *35*(6), 1167-1175.

Hamilton, A. F. de C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *26*(4), 1133-1137. doi:10.1523/JNEUROSCI.4551-05.2006

Hamilton, A. F. de C., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *18*(5), 1160-1168. doi:10.1093/cercor/bhm150

Harries, M.H., & Perrett, D.I. (1991). Visual processing of faces in temporal cortex: physiological evidence for a modular organization and possible anatomical correlates. *Journal of Cognitive Neuroscience, 3*, 9–24.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of cognitive neuroscience*, *21*(7), 1229-1243. doi:10.1162/jocn.2009.21189

Horner, V., Whiten, A., Flynn, E., & de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13878-13883. doi:10.1073/pnas.0606015103

Huffman, M.A. (1996). The study of nonhuman primate culture in Japan. In Heyes CM, Galef BG (Eds). Social learning in animals: the roots of culture (pp-267-289). San Diego, CA: Academic Press.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science (New York, N.Y.), 286*(5449), 2526-2528.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., Rizzolatti, G.(2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3(3):e79.*

Jacquet, P.O., Tessari, A., Binkofski, F., & Borghi, A.M. (2012a). Can object affordances impact on human social learning of tool use? *Behavioural and Brain Sciences, in press.*

Jacquet, P.O., Chambon, V., Borghi, A.M., Tessari, A. (2012b). Object Affordances Tune Observers' Prior Expectations About Tool-Use Behaviors, PLoS One, *in press*

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in cognitive sciences*, *9*(1), 21-25. doi:10.1016/j.tics.2004.11.003

Jeannerod, M. (1995). The representing brain: Neural correlates of motor intentions and imagery. *Behavioural and Brain Sciences, 17(02)*, 187-202. doi: http://dx.doi.org/10.1017/S0140525X00034026

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, *14*(1 Pt 2), S103-109. doi:10.1006/nimg.2001.0832

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., Voss, H. U., et al. (2011). Behavioral and neural properties of social reinforcement learning. *The Journal of neuroscience: the official journal of the Society for Neuroscience, 31*(37), 13039-13045. doi:10.1523/JNEUROSCI.2972-11.2011

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005) Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Developmen*t, 20, 303–320.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual review of psychology*, *55*, 271-304. doi:10.1146/annurev.psych.55.090902.142005

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in cognitive sciences*, *11*(5), 194-196. doi:10.1016/j.tics.2007.02.002

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature reviews. Neuroscience*, *11*(6), 417-428. doi:10.1038/nrn2833

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357. doi:10.1016/j.tics.2011.06.005

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). The mirror-neuron system: a Bayesian perspective. *Neuroreport*, *18*(6), 619-623. doi:10.1097/WNR.0b013e3281139ed0

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166. doi:10.1007/s10339-007-0170-2

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition, 83*(2), 35-42.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, *61*(1), 140-151. doi:10.1016/j.neuron.2008.11.027

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., et al. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20084-20089. doi:10.1073/pnas.0710103104

Khün, S., Haggard, P., Brass, M. (2009). Intentional inhibition: how the "veto-area" exerts control. Human Brain Mapping, 30(9), 2834-28-43.

Lago, A., & Fernandez-del-Olmo, M. (2011). Movement observation specifies motor programs activated by the action observed objective. *Neuroscience letters*, *493*(3), 102-106. doi:10.1016/j.neulet.2011.02.014

Laland, K.N., & Galef, B.G. (2009). The question of animal culture. Cambridge, MA: Harvard University Press.

Laplace, P.S. (2000). Theorie analytique des probabilités. (1814). Broché ed: Paris.

Leslie, A. M. (1987). Pretense and representation: the origins of “theory of mind”. *Psychological Review*, 94, 412-42.

Li, L., Miller, E.K., Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *The Journal of Neurophysiology*, *69,* 1918-1929.

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological science*, *16*(8), 601-608. doi:10.1111/j.1467-9280.2005.01582.x

Luppino, G., Murata, A., Govoni, P., & Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *128*(1-2), 181-187.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1-9.

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., Lusseau, D., et al. (2007). Cetaceans have complex brains for complex cognition. *PLoS biology*, *5*(5), e139. doi:10.1371/journal.pbio.0050139

Metzinger, T., & Gallese, V. (2003). The emergence of a shared action ontology: building blocks for a theory. *Consciousness and cognition*, *12*(4), 549-571.

Molnar-Szakacs, I., Kaplan, J., Greenfield, P.M., Iacoboni, M. (2006). Observing complex action sequences: The role of the fronto-parietal mirror neuron system. *NeuroImage, 33*, 923-935.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current biology: CB*, *20*(8), 750-756. doi:10.1016/j.cub.2010.02.045

Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., & Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of neurophysiology*, *78*(4), 2226-2230.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of neurophysiology*, *83*(5), 2580-2601.

Newell, B. R. (2005). Re-visions of rationality? *Trends in cognitive sciences*, *9*(1), 11-15. doi:10.1016/j.tics.2004.11.005

Norman, D.A. (1988). The Psychology of Everyday Things. New York, NY: Basic Books.

Oram, M.W., & Perrett, D.I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to biological motion stimuli. *Journal of Cognitive Neuroscience, 6*, 99–116

Osiurak, F., Jarry, C., & Le Gall, D. (2010). Grasping the affordances, understanding the reasoning: toward a dialectical theory of human tool use. *Psychological review*, *117*(2), 517-540. doi:10.1037/a0019004

Peterson, C.R. and Beach, L.R. (1967) Man as an intuitive statistician. *Psychological Bulletin, 68*, 29–46

Petrie, M. (1994). Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature*, *371*: 598-599. doi:10.1038/371598a0

Pierno, A.C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., Castiello, U. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cerebral Cortex, 19(2)*, 367-374.

Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, *259*(2), 76-83.

Reber, A.S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Reber, A.S. (1993). Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious. Oxford University Press.

Rendell, L., Fogarty, L., Hoppitt, W. J. E., Morgan, T. J. H., Webster, M. M., & Laland, K. N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in cognitive sciences*, *15*(2), 68-76. doi:10.1016/j.tics.2010.12.002

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*(4), 1694-1703. doi:10.1016/j.neuroimage.2004.04.015

Rizzolatti G., *et al.* (2000). Cortical mechanism subserving object grasping and action recognition: a new view on the cortical motor functions. In M., Gazzaniga, Ed. *The New Cognitive Neurosciences*, pp. 539-552.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research*, *71*, 491-507.

Rizzolatti, G, & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194.

Rizzolatti, G, Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *71*(3), 491-507.

Rizzolatti, G, Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain research. Cognitive brain research*, *3*(2), 131-141.

Rizzolatti, G, Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews. Neuroscience*, *2*(9), 661-670. doi:10.1038/35090060

Rizzolatti, Giacomo, & Craighero, L. (2004b). The mirror-neuron system. *Annual review of neuroscience*, *27*, 169-192. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, Giacomo, & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews. Neuroscience*, *11*(4), 264-274. doi:10.1038/nrn2805

Romani, M., Cesari, P., Aglioti, S. (2003).Riconoscere ed eseguire un movimento: sapere e saper fare in giocatori professionisti di pallacanestro, studi psicofisici e di stimolazione magnetica transcranica (Movement recognition and movement execution in basketball players: knowing and knowing by doing). *Movimento*, 19 (2), 58–61.

Rosenbaum, D. A., van Heugten, C. M., & Caldwell, G. E. (1996). From cognition to biomechanics and back: the end-state comfort effect and the middle-is-faster effect. *Acta psychologica*, *94*(1), 59-85.

Rosenbaum, D. A., Vaughan, J., Barnes, H. J., & Jorgensen, M. J. (1992). Time course of movement planning: selection of handgrips for object manipulation. *Journal of experimental psychology. Learning, memory, and cognition*, *18*(5), 1058-1073.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926-1928.

Sartori, L., Straulino, E., & Castiello, U. (2011). How objects are grasped: the interplay between affordances and end-goals. *PloS one*, *6*(9), e25203. doi:10.1371/journal.pone.0025203

Saxe, R. (2005). Against simulation: the argument from error. *Trends in cognitive sciences*, *9*(4), 174-179. doi:10.1016/j.tics.2005.01.012

Schmidt, D. M., George, J. S., & Wood, C. C. (1999). Bayesian inference applied to the electromagnetic inverse problem. *Human brain mapping*, *7*(3), 195-212.

Seed, A., & Byrne, R. (2010). Animal tool-use. *Current biology: CB*, *20*(23), R1032-1039. doi:10.1016/j.cub.2010.09.042

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *The Journal of comparative neurology*, *343*(3), 445-463. doi:10.1002/cne.903430308

Silvanto, J., Muggleton, N. G., Cowey, A., & Walsh, V. (2007). Neural adaptation reveals state-dependent effects of transcranial magnetic stimulation. *The European journal of neuroscience*, *25*(6), 1874-1881. doi:10.1111/j.1460-9568.2007.05440.x

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in cognitive sciences*, *12*(12), 447-454. doi:10.1016/j.tics.2008.09.004

Silvanto, J., Pascual-Leone, A. (2008). State-dependency of transcranial magnetic stimulation. *Brain Topography*, *21*, 1-10.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, *41*, 1-19. doi:10.1146/annurev.ps.41.020190.000245

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In Carrol JS and Payne JW, Eds. *Cognition and Social Behavior*. pp.165-184, Erlbaum.

Sodian, B., Schoeppner, B. & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development*, *27*, 31–41.

Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059-1069. doi:10.1016/j.cognition.2007.10.002

Sperber D. (1996). Explaining culture. A naturalistic approach. Oxford: Blackwell.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in cognitive sciences*, *8*(1), 40-46.

St Amant, R., Horton, T.E. (2008). Revisiting the definition of animal tool use. *Animal Behaviour*, *75*: 1199-1208.

Stefan, K, Classen, J., Celnik, P., & Cohen, L. G. (2008). Concurrent action observation modulates practice-induced motor memory formation. *The European journal of neuroscience*, *27*(3), 730-738. doi:10.1111/j.1460-9568.2008.06035.x

Stefan, Katja, Cohen, L. G., Duque, J., Mazzocchio, R., Celnik, P., Sawaki, L., Ungerleider, L., et al. (2005). Formation of a motor memory by action observation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *25*(41), 9339-9346. doi:10.1523/JNEUROSCI.2282-05.2005

Sterelny, K. (2003a). Thought in a hostile world. New York, NY: Blackwell.

Sterelny, K. (2003b). Cognitive load and human decision, or, three ways of rolling the rock uphill. In: Carruthers P, Laurence S, Stich S, Eds. *The innate mind: Vol2, Culture and cognition*. Oxford, UK: Oxford University Press. pp 218-233

Sterelny, K. (2009). Peacekeeping in the culture wars. In Laland, K.N., & Galef, B.G, Eds. *The question of animal culture.* Cambridge, MA: Harvard University Press.

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*(1585), 75-87. doi:10.1098/rstb.2011.0099

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*(9), 403-409. doi:10.1016/j.tics.2009.06.003

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science (New York, N.Y.)*, *331*(6022), 1279-1285. doi:10.1126/science.1192788

Tennie, C., Hedwig, D., Call, J., & Tomasello, M. (2008). An experimental study of nettle feeding in captive gorillas. *American journal of primatology*, *70*(6), 584-593. doi:10.1002/ajp.20532

Tomasello, M. (2009). The cultural origins of human cognition. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and brain sciences*, *28*(5), 675-691; discussion 691-735. doi:10.1017/S0140525X05000129

Trivers, R.L. (1971). The evolution of reciprocal altruism. The quaterly review of biology, 46(1), 35-37.

Tunik, E., Rice, N.J., Hamilton, A., & Grafton, S.T. (2007). Beyond grasping: representation of action in the human anterior intraparietal sulcus. *NeuroImage*, 36(2), 77-86.

Vaesen, K. (2012). The cognitive bases of human tool-use. *Behavioural and Brain Sciences.*

van Elswijk, G., Kleine, B. U., Overeem, S., & Stegeman, D. F. (2007). Expectancy induces dynamic modulation of corticospinal excitability. *Journal of cognitive neuroscience*, *19*(1), 121-131. doi:10.1162/jocn.2007.19.1.121

Visalberghi, E., Addessi, E., Truppa, V., Spagnoletti, N., Ottoni, E., Izar, P., & Fragaszy, D. (2009). Selection of effective stone tools by wild bearded capuchin monkeys. *Current biology: CB*, *19*(3), 213-217. doi:10.1016/j.cub.2008.11.064

Voelkl, B., & Huber, L. (2007). Imitation as faithful copying of a novel technique in marmoset monkeys. *PloS one*, *2*(7), e611. doi:10.1371/journal.pone.0000611

Wagner, L., & Carey, S. (2005). 12-month-old infants represent probable ending of motion events. *Infancy, 7,* 73–83.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature reviews. Neuroscience*, *1*(1), 73-79. doi:10.1038/35036239

Vainio, L., Ellis, R., & Tucker, M. (2007). The role of visuospatial attention in action priming. *The Quarterly Journal of Experimental Psychology, 60*, 241–261.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, *48*(3), 564-584. doi:10.1016/j.neuroimage.2009.06.009

Weiss, D. J., Wark, J. D., & Rosenbaum, D. A. (2007). Monkey see, monkey plan, monkey do: the end-state comfort effect in cotton-top tamarins (Saguinus oedipus). *Psychological science*, *18*(12), 1063-1068. doi:10.1111/j.1467-9280.2007.02026.x

Whiten, A, Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C. E., et al. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682-685. doi:10.1038/21415

Whiten, Andrew. (2011a). The scope of culture in chimpanzees, humans and ancestral apes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *366*(1567), 997-1007. doi:10.1098/rstb.2010.0334

Whiten, Andrew, Horner, V., & de Waal, F. B. M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, *437*(7059), 737-740. doi:10.1038/nature04047

Whiten, Andrew, Horner, V., Litchfield, C. A., & Marshall-Pescini, S. (2004). How do apes ape? *Learning & behavior*, *32*(1), 36-52.

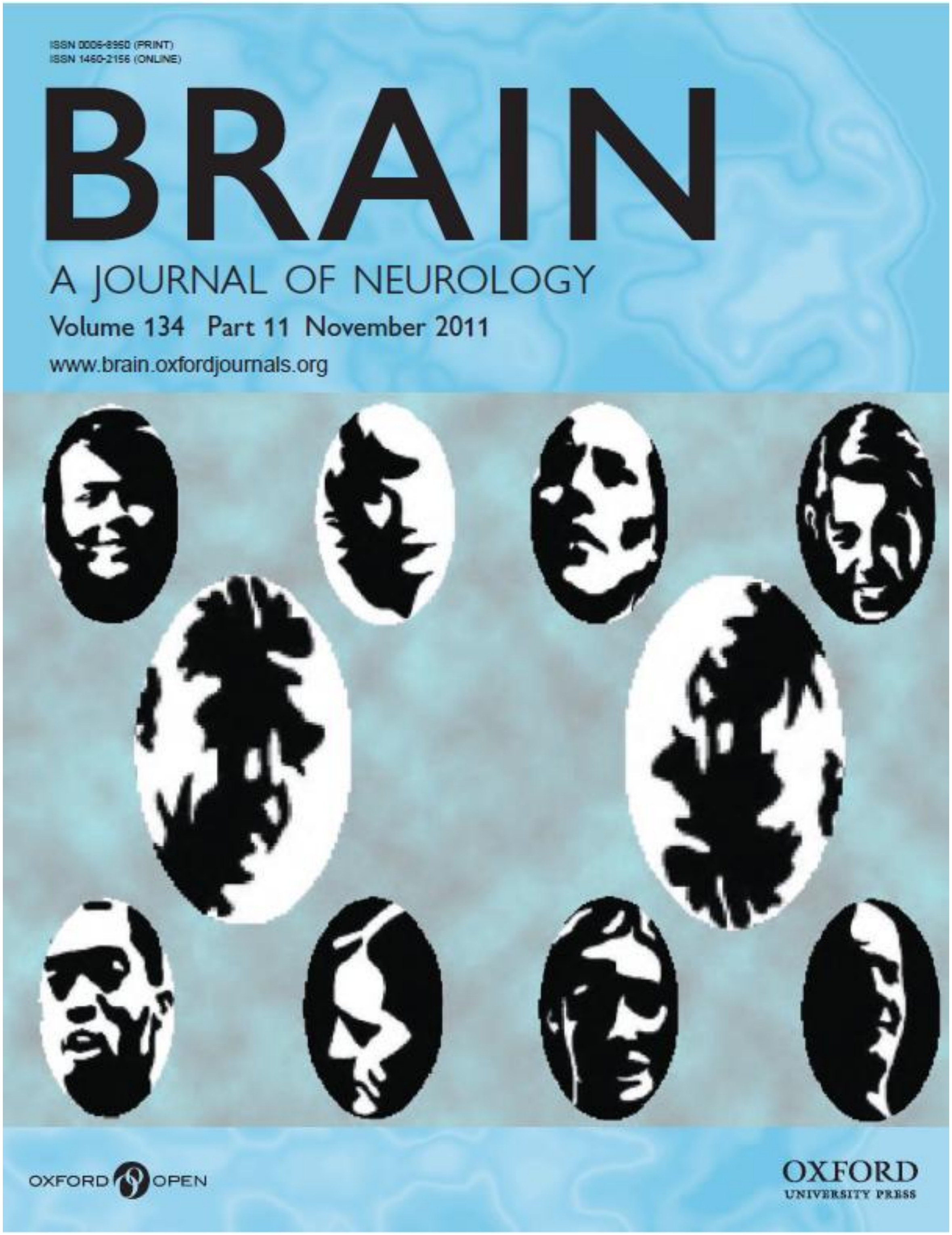
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 593-602. doi:10.1098/rstb.2002.1238

Wu, R., Gopnik, A., Richardson, D.C., & Kirkham, N.Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology, 47*(5), 1220-1229.

APPENDIX I

Other articles published or under revision in international peer reviews



ISSN 0006-8950 (PRINT)
ISSN 1460-2156 (ONLINE)

BRAIN

A JOURNAL OF NEUROLOGY

Volume 134 Part 11 November 2011
www.brain.oxfordjournals.org

OXFORD OPEN

OXFORD
UNIVERSITY PRESS

doi:10.1093/brain/awr306 Brain 2011: 134; 3728–3741 | 3728

BRAIN
A JOURNAL OF NEUROLOGY

Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia

Valerian Chambon,[1] Elisabeth Pacherie,[2,*] Guillaume Barbalat,[3,4,*] Pierre Jacquet,[1]
Nicolas Franck[1,4] and Chlöé Farrer[5,6]

1 Centre de Neuroscience Cognitive, Université de Lyon, CNRS, France
2 Institut Jean Nicod, EHESS, DEC-ENS, CNRS, Paris, France
3 Laboratoire Langage, Cerveau, Cognition, Université de Lyon, CNRS, France
4 Centre Hospitalier le Vinatier, Lyon, France
5 Centre de Recherche Cerveau et Cognition, Université de Toulouse, UPS-CNRS, Toulouse, France
6 Faculté de Médecine de Rangueil, Toulouse, France

*These authors contributed equally to this work.

Correspondence to: Valerian Chambon,
Centre de Neuroscience Cognitive – UMR 5229 CNRS;
67, bd Pinel, 69675 BRON Cedex,
France
E-mail: valerian.chambon@isc.cnrs.fr

An impaired ability to appreciate other people's mental states is a well-established and stable cognitive deficit in schizophrenia, which might explain some aspects of patients' social dysfunction. Yet, despite a wealth of literature on this topic, the basic mechanisms underlying these impairments are still poorly understood, and their links with the clinical dimensions of schizophrenia remain unclear. The present study aimed to investigate the extent to which patients' impaired ability to appreciate other people's intentions (known as mentalizing) may be accounted for by abnormal interaction between the two types of information that contribute to this ability: (i) the sensory evidence conveyed by movement kinematics; and (ii) the observer's prior expectations. We hypothesized that this is not a generalized impairment, but one confined to certain types of intentions. To test this assumption, we designed four tasks in which participants were required to infer either: (i) basic intentions (i.e. the simple goal of a motor act); (ii) superordinate intentions (i.e. the general goal of a sequence of motor acts); (iii) social basic; or (iv) social superordinate intentions (i.e. simple or general goals achieved within the context of a reciprocal interaction). In each of these tasks, both prior expectations and sensory information were manipulated. We found that patients correctly inferred non-social, basic intentions, but experienced difficulties when inferring non-social superordinate intentions and both basic and superordinate social intentions. These poor performances were associated with two abnormal patterns of interaction between prior expectations and sensory evidence. In the non-social superordinate condition, patients relied heavily on their prior expectations, while disregarding sensory evidence. This pattern of interaction predicted the severity of 'positive' symptoms. Social conditions prompted exactly the opposite pattern of interaction: patients exhibited weaker dependence on prior expectations while relying strongly on sensory evidence, and this predicted the severity of 'negative' symptoms. We suggest both these patterns can be accounted for by a disturbance in the Bayesian inferential mechanism that integrates sensory evidence (conveyed by movement kinematics) into prior beliefs (about others' mental states and attitudes) to produce accurate inferences about other people's intentions.

Received May 24, 2011. Revised August 25, 2011. Accepted September 27, 2011. Advance Access publication November 21, 2011

Keywords: schizophrenia; mentalizing; sensory evidence; prior expectations; Bayesian integration
Abbreviation: TFT = tit-for-tat

Introduction

One of the most disabling clinical features of schizophrenia is poor social functioning, reflecting impairments in interpersonal communication and relationships (see Corcoran, 2001, for review). Many authors have proposed that some aspects of patients' social dysfunction are a consequence of a deficit in mentalizing, defined as the cognitive ability to attribute mental states (such as intentions) to others and explain and predict their behaviour on that basis (Frith, 2004; Harrington *et al.*, 2005; Sprong *et al.*, 2007). Extensive research over the last two decades has provided robust evidence for the presence of a stable mentalizing impairment in schizophrenia (Sprong *et al.*, 2007; Bora *et al.*, 2009). However, both the nature and the extent of this impairment remain widely debated, owing to its extreme heterogeneity among clinical subgroups of schizophrenia (McCabe *et al.*, 2004; Harrington *et al.*, 2005; Bara *et al.*, 2011).

It has been suggested that inconsistent results in the literature may be the consequence of the great variety of tasks used, both in terms of stimulus type (verbal versus iconographic) and complexity (Walter *et al.*, 2009). Crucially, the heterogeneity of the data could also result from a lack of control over the variable under examination. Indeed, 'intention' is a term embracing various subtypes, the content of which can vary along two main dimensions: the scope and the target (Chambon *et al.*, 2011). The 'scope dimension' refers to the complexity of the intended goal, and differentiates 'basic intentions' directed at simple motor goals (e.g. grasping an object) from 'superordinate intentions' directed at somewhat more complex goals (e.g. quenching one's thirst), the achievement of which typically involves the completion of a number of subgoals (e.g. grasping a glass, opening a tap, filling the glass, closing the tap, etc.) (Pacherie, 2000, 2008). On the 'target dimension', 'non-social intentions' directed at an object can be distinguished from 'social intentions' directed at a third party (Blakemore and Frith, 2004; Ciaramidaro *et al.*, 2007). The ability to appreciate other people's intentions thus refers to separate processes that could be differentially recruited depending on the scope and/or the target of the intention being considered. As such, one cannot preclude the possibility that patients may show impaired understanding of one particular type of intention while the appreciation of other intention types is spared.

So far, few studies have directly tested patients' abilities to appreciate distinct types of intention within the same experimental settings, or using the same material across conditions. One study found that disorganized patients were impaired at evaluating superordinate intentions but not basic intentions (Zalla *et al.*, 2006). Another recent study suggested that patients may not be impaired in appreciating actions directed at inanimate objects, but specifically in inferring intentions achieved within the context of social interaction (Walter *et al.*, 2009). Disentangling this confusing array of findings requires investigating patients' mentalizing abilities at a more fine-grained level of functioning. That is, not only by assessing patients' raw performances in intention recognition tasks, but also by further exploring how individuals with schizophrenia deal with the information that usually contributes to such recognition.

Attributing intentions to an observed agent can be described as a Bayesian inference drawing upon two distinctive types of information: (i) the 'sensory evidence' available from the action scene (derived from the agent's movement kinematics); and (ii) the observer's 'prior expectations' about which intention is the most likely cause of what is observed, given past experience (Baker *et al.*, 2006, 2009; Griffiths *et al.*, 2008). It has been shown that intention inference is contingent upon an adaptive interplay between these two sources of information, with participants tending to rely progressively more on their prior expectations as the reliability of sensory evidence decreases, and vice versa. Crucially, this interaction has also been found to vary according to the 'type' of intention to be inferred, with participant's prior experience gaining priority over perceptual evidence when inferring intentions from within a social context rather than in isolation (Chambon *et al.*, 2011).

Building on these previous findings, we hypothesized that patients' heterogeneous mentalizing abilities could be accounted for by an abnormal weighting of these two classes of information (prior knowledge and sensory evidence), which in turn might depend on the specific dimensions (i.e. the scope and target) of the intention being considered. This assumption echoes Fletcher and Frith's (2009) suggestion that both the aberrant perceptions (hallucinations) and beliefs (delusions) of schizophrenia might be caused by an abnormality in the brain's inferencing mechanisms, resulting in a diminished ability to integrate new experiences (e.g. sensory evidence) with stored knowledge based on previous experiences (e.g. prior knowledge; Hemsley, 2005). Critically, disturbance of this (Bayesian) inferential mechanism could be a good predictor of the severity of schizophrenia symptoms. For example, the mentalizing profile of patients with positive symptoms might be characterized by a tendency to give excessive credit to endogenous, self-generated information (e.g. prior expectations of how people are supposed to behave under some circumstances), whereas patients with negative symptoms might display a stimuli-induced mentalizing style that may be accounted for by an exaggerated tendency to focus on directly observable, external information, rather than inner experiences (Frith, 1994; Taylor, 1994).

In the present study, we directly tested the above assumption by assessing patients' understanding of the basic or superordinate intentions of an agent performing an action in either isolation, or within the context of social reciprocation. Both sensory and prior information were manipulated by: (i) varying the completeness of action sequences; and (ii) selectively increasing the probability of a particular intention occurring within the sequence, at the expense of competing intention types. We then looked at (i) whether patients' performances on each intention inference task may be accounted for by an abnormal dependence on prior knowledge and/or sensory evidence, and (ii) whether this abnormal dependence—if observed—correlated with the scale for the assessment of positive (Andreasen, 1984), negative (Andreasen, 1983) or disorganization symptoms of schizophrenia.

Materials and methods

Participants

All patients fulfilled DSM-IV criteria of schizophrenia (American Psychiatric Association, 1994) with no other psychiatric diagnosis on DSM-IV Axis I. Exclusion criteria included history of neurological illness or trauma, alcohol or drug dependence according to DSM-IV criteria, analphabetism and being >60 years of age. All patients were receiving antipsychotic medication and were clinically stable at the time of testing. Comparison participants reported no psychiatric problems (Table 1), and were systematically matched with patients for age, handedness (Oldfield, 1971) and years of education (Table 1). All participants reported normal or corrected-to-normal visual acuity. After receiving a complete description of the study, written informed consent was obtained according to the Declaration of Helsinki. This research was approved by the local Ethical Committee (B80631-60) and all participants received 10 euros for taking part.

Four distinct groups of controls ($n = 30$ for each group) and patients ($n = 20$ for each group) performed the four distinct tasks. Individuals with schizophrenia were selected to obtain four groups of patients matched for the severity of negative (scale for the assessment of negative symptoms; Andreasen, 1983), positive (scale for the assessment of positive symptoms; Andreasen, 1984) and disorganization symptoms (Table 1). The disorganization score was computed by summing the following subscores: bizarre behaviour, positive formal thought disorder (from the scale for the assessment of positive symptoms), alogia and inappropriate affect (from the scale for the assessment of negative symptoms). These items have been shown to constitute regular and fundamental components of the disorganization dimension (Hardy-Bayle *et al.*, 2003). In the social basic task, one patient was excluded because of poor performance [i.e. >2 standard deviations (SDs)] from the group mean).

Common procedure in the four tasks

In each task, participants were instructed to infer the intention of an actor manipulating non-meaningful objects. The specific contributions of sensory evidence and prior knowledge to the intentional inference were manipulated by varying the amount of visual information (i.e. the completeness of action sequences) and the probability of occurrence associated with each different intention, respectively [see Chambon *et al.* (2011) for detailed descriptions of the video clips used in each task].

Each task consisted of two experimental sessions. First, a baseline session, characterized by a flat (unbiased) probability distribution, in which all intentions had the same probability of occurrence across trials. Secondly, a bias session, in which prior knowledge was manipulated by increasing the probability of one intention (the 'likely' intention, 55% of the trials) to the detriment of the others ('unlikely' intentions, 22% each), resulting in biasing participants towards the likely intention. This bias was randomly assigned so that each intention was equally biased across participants.

Table 1 Clinical and demographic characteristics

Characteristics	Age (years)	Education (years)	Handedness	Duration of illness	SANS score	SAPS score	Disorganization score[a]
Experiment							
Non-social basic							
Healthy ($n = 30$)	35.1 (7.5)	11.9 (2)	0.87 (0.14)				
Patients ($n = 20$)	34 (9.3)	11.1 (1.7)	0.83 (0.16)	10.3 (7.5)	40.5 (15)	31.9 (23.5)	16.1 (12.9)
P-value	0.65	0.12	0.37				
Non-social superordinate							
Healthy ($n = 30$)	36.5 (8.9)	12.1 (1.5)	0.81 (0.17)				
Patients ($n = 20$)	34.6 (8.8)	11.6 (1.8)	0.78 (0.17)	12.3 (8.1)	43.2 (21.6)	29.9 (15.6)	12.9 (5.3)
P-value	0.46	0.26	0.61				
Social basic							
Healthy ($n = 30$)	34.2 (10.5)	11.4 (1.8)	0.82 (0.14)				
Patients ($n = 19$)	35.2 (9)	11.2 (1.7)	0.79 (0.19)	11 (8.4)	44 (24.1)	28.5 (22.3)	14.5 (12.8)
P-value	0.74	0.61	0.57				
Social superordinate							
Healthy ($n = 30$)	35.4 (8.8)	12.3 (1.9)	0.85 (0.13)				
Patients ($n = 20$)	33.8 (10)	11.7	0.81 (0.18)	11.9 (8.6)	44.8 (23.9)	29.4 (15.1)	11.5 (6.9)
P-value	0.56	0.27 (1.4)	0.48				
				all $P > 0.43$	all $P > 0.5$	all $P > 0.64$	all $P > 0.14$

[a] Sum of the scores for bizarre behaviour, positive formal thought disorder from the SAPS, and alogia and inappropriate affect from the SANS.
SANS = Scale for the Assessment for the Negative Symptoms; SAPS = Scale for the Assessment of Positive Symptoms. Data are mean (SD).

The amount of visual information was manipulated by varying the duration of the video clips. Actions were thus either presented with a very high (1880 ms after movement onset), high (1640 ms), moderate (1560 ms), or low (1480 ms) amount of visual information [see Chambon *et al.* (2011) for the selection and control of these amounts].

The baseline and the bias sessions were composed of two types of interleaved blocks: 'overt' blocks, in which the actions were shown with a very high amount of visual information (1880 ms) to allow participants to clearly distinguish the different intentions, and 'covert' blocks, in which actions were of varying durations (1480, 1560 or 1640 ms) (Fig. 1). The overt blocks were used to bias participants in favour of one particular intention (i.e. the *likely* intention), whereas the covert blocks were used to test the effect of the bias on action sequences shown with varying amounts of visual information.

Each experimental sequence (one overt block followed by one covert block) was repeated nine times over each session. The order of trials was randomized and varied between participants. Furthermore, each clip was presented only once to prevent any influence of memorized kinematic parameters on participants' performances.

All clips were filmed using a digital camera (Sony®- HDR-SR7) and were acquired and tailored using the software Adobe Premiere®. They were presented on a computer monitor (IIYAMA® 19') at a distance of 60 cm from the participant. Finally, prior to each task, a training session was conducted with distinct clips from those used in the experimental sessions.

Non-social tasks

In both the non-social basic and the non-social superordinate tasks, video clips depicted a single actor manipulating (rotating, lifting or transporting) rectangular cubes. The cubes were of similar size (3 × 6 cm) and orientation, and placed at an equal distance (16.8 cm) from the starting position of the actor's hand (Fig. 2A and B).

Non-social basic task

In the non-social basic task, participants were first required to observe one incomplete manipulation of a single cube (lasting for 1480, 1560 or 1640 ms after movement onset). A response screen representing the first letter of each possible non-social, basic intention (to transport, lift or rotate) then appeared for 2500 ms, during which participants had to press the keyboard button corresponding to the intention inferred (transport, lift or rotate) as quickly and accurately as possible. In the bias session, the non-social basic intention for which the probability of occurrence was increased (i.e. the likely intention) was counterbalanced across participants.

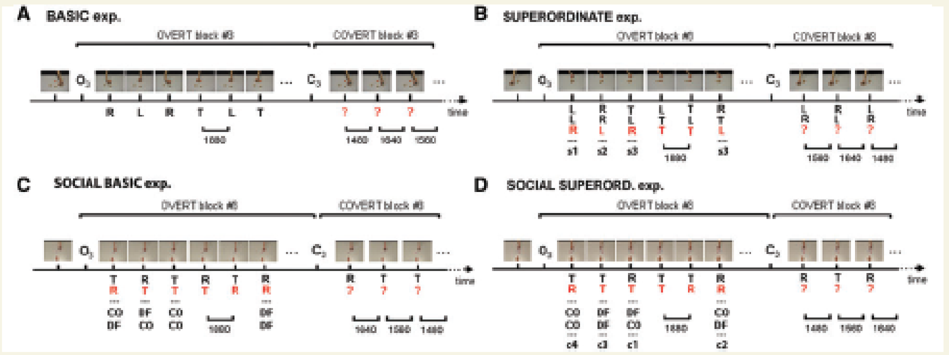


Figure 1 Task design. Examples of the typical experimental sequence (one overt block followed by one covert block) used in both the baseline and the bias sessions. Overt blocks (O): 18 movies with a very high, constant amount of visual information (1880 ms). Covert blocks (C): nine movies with three different amounts of visual information (1480, 1560 and 1640 ms). In the four tasks, the probability of all intentions was held constant across the block, except in the overt blocks of the bias session, where one particular intention had a greater probability of occurring than the others. (A) In the basic task, subjects had to identify a single intended action (L = lift; R = rotate; T = transport). (B) In the superordinate task, subjects had to identify the final intended action (indicated by a red letter) of an action sequence leading to shapes 1, 2 or 3 (s1 = shape 1, etc.). (C) In the social basic task, subjects had to identify the intended action of the second player (red letter). (D) In the social superordinate task, subjects had to identify the intended action of the second player (red letter) leading to configurations 1, 2, 3 or 4 (c1 = configuration 1, etc.). In both the social basic and social superordinate tasks, the action or the configuration achieved by each player indicated either a cooperative or a defective strategy (CO = cooperate; DF = defect). In each experiment, a probabilistic bias was assigned to one particular action (basic), shape (superordinate) or strategy (social). The red question mark indicates the action for which the amount of visual information varied (basic: a single action; superordinate: the last action of the sequence).

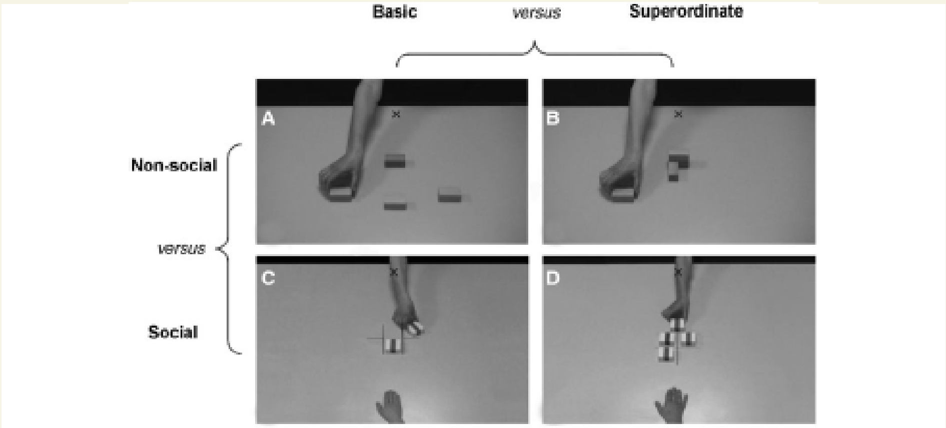


Figure 2 Examples of stimuli for each of the four tasks. (**A**) Non-social basic intention task; (**B**) non-social superordinate intention task; (**C**) social basic intention task; and (**D**) social superordinate intention task. The black cross indicates the starting position of the hand.

Non-social superordinate task

Video clips showed a sequence of three actions (e.g. to transport, to rotate or to lift a cube) leading to the construction of one out of three possible non-meaningful shapes (s1, s2 or s3; Fig. 1B). Each sequence was therefore characterized by a superordinate intention to build one of these three final shapes. The duration of the video sequences was varied (lasting 1480, 1560 or 1640 ms after movement onset) so that the last action was rendered incomplete. Participants were instructed to infer the superordinate intention and to give a response indicating the nature of the last, incomplete action in the sequence by pressing the corresponding keyboard button as quickly and accurately as possible.

Crucially, to ensure that participants were biased towards the superordinate intention itself and not merely towards the final action, commutative (i.e. interchangeable) sequences were used so that each shape could be constructed from multiple, distinct sequences of actions. Sequences shown in the covert blocks were thus distinct from those used in the overt blocks (e.g. the shape s1 could be obtained from the sequence 'lift–lift–rotate' in an overt block, but from the sequence 'lift–rotate–lift' in a covert block). In the bias session, the probability of building one of the three final shapes was increased at the expense of the other two, whilst keeping the probability of each simple action occurring during shape-building equal. The specific shape that was biased was counterbalanced across participants.

Social intention tasks

In the two social tasks (social basic and social superordinate), participants were instructed to infer whether a social intention was of either a cooperative or defective nature. They observed two players engaged in a social game, in which they either cooperated by coordinating their actions in order to achieve a shared goal, or defected by refusing to coordinate their actions. One after the other, the two players either transported the cube closest to them towards the centre of the board, or rotated it so that it remained at the same place (Fig. 2C and D). The first player's action was always shown entirely to the participants, while the second player's action was made incomplete by varying its duration across the trials (1480, 1560 or 1640 ms after its onset). Participants had to infer the nature of the second player's social intention (i.e. cooperative or defective). To do so, they were instructed to give a response about the nature of the incomplete action (i.e. to rotate or to transport) which unambiguously denoted the social intention, by pressing the corresponding button as quickly and accurately as possible (R for rotate, T for transport).

In the video clips, the second player's social intention either differed from that of the first player (i.e. the first player defected and the second cooperated, or the first player cooperated while the second defected) or it mirrored the first player's intention (i.e. both players cooperated or defected). This second type of response strategy is known as a 'tit-for-tat' (TFT) strategy. In situations of iterative cooperation, a TFT strategy is known to frequently be a more intuitive and successful strategy than alternative ones, such as 'always cooperating', 'always defecting' or 'acting randomly' (Axelrod, 1997; André and Day, 2007; Chambon *et al.*, 2011). We thus chose to experimentally strengthen this existing *a priori* bias by increasing the probability that the second player adopts a TFT strategy, i.e. uses a strategy that mirrors their opponent's. In the bias session, the probability that the second player responded TFT was therefore increased so that, on average, he was more likely to cooperate (rather than defect) if the first player had previously cooperated, and to defect (rather than cooperate) if the first player had previously defected. In the baseline session, however, the probability of a TFT response was equal to that of responses using alternative strategies (i.e. cooperation in response to defection, or defection in response to cooperation).

Biasing the second player's strategy in this way ensured that participants paid attention to the whole action sequence, since to successfully predict the intentions of a player using a TFT strategy it is essential to take into account what the first player has done. Furthermore, using a TFT bias also prevented participants from giving stereotyped responses (e.g. always responding 'cooperate' or 'defect').

Social basic task

In the social basic task, participants were required to infer a social (defective or cooperative) intention that was denoted by the second player's action. This action consisted of either transporting a cube object (printed with a red or a blue line) towards the middle of a grid (termed 'bank') or rotating it so that it remained in its original location.

Social superordinate task

In the social superordinate task, the social intention inferred was achieved by the sequence of both players' actions and therefore corresponded to a final configuration of cubes (Fig. 1D). Players acted in turn with the goal to vertically align three cubes among the four available ones (one cube was printed with a blue line; the other three with a red line). The individual goals of the first and the second players were to align the three cubes printed with a red line (irrespective of the orientation of the lines) or to align three cubes with the same line orientation (irrespective of line colour), respectively. Combining both possible strategies for each player resulted in four possible final configurations of the cubes: both players defected, each preventing the other from achieving his goal (Configuration 4: no alignment); both players cooperated, in order to achieve both of their goals (Configuration 3: cubes were aligned according to both their colour and line orientation); the first player defected, preventing the second from achieving his goal, whilst the second cooperated, helping the first achieve his goal (Configuration 1: cubes were aligned according to their colour); or finally, the first player cooperated, helping the second achieve his goal, whilst the second defected preventing the first from achieving his goal (Configuration 2: cubes were aligned according to line orientation) (Fig. 1D).

As in the non-social superordinate task, commutative sequences were used so that each configuration could be obtained from distinct sequences of actions, ensuring that the second player's intention (e.g. playing TFT) could not be predicted from his single action (e.g. to rotate, or to transport) but only from the entire sequence of actions. Furthermore, the overall probabilities of each strategy (cooperative or defective) and of each single action (to rotate or to transport) were kept equal across the blocks.

Data analyses

Hits and reaction times

In the overt blocks of the bias and baseline sessions, patients' and controls' percentage of correct responses (hits) were compared using two-sample *t*-tests. These analyses were performed to ensure that both groups were equally successful in integrating the flat (baseline session) and biased (bias session) probability distributions associated with each intention.

Note that intentions were equally probable in the baseline session. We therefore referred to as 'future' likely (f-likely) and 'future' unlikely (f-unlikely) those intentions whose probability was increased (likely intention), or decreased (unlikely intention), in the subsequent bias session.

In the covert blocks of the baseline and bias sessions, hits and reaction times were analysed independently using $2 \times 2 \times 3$ mixed-model, repeated-measures ANOVAs with group (controls versus patients) as a between-subjects factor, and intention (f-likely versus f-unlikely intentions) or bias (likely versus unlikely intentions), and amounts of visual information (low, moderate and high) as within-subjects factors. *Post hoc* Fisher tests were then performed to identify differences between conditions.

Whenever the variance structure did not conform to the requirements for parametric analyses, logarithmic transformations were used to obtain the required conformity. Analyses were performed using the statistical software Statistica 7 (www.statsoft.com).

Bias effect

To assess whether the assignment of a bias differently affected the performance of patients compared with controls across the four types of intention, a score reflecting the 'bias effect' was calculated for each subject, in each task. This score was obtained by subtracting the number of correct responses for the likely intention from those of the unlikely ones, in the covert blocks of the bias session. We then performed a $3 \times 4 \times 2$ repeated-measures ANOVA with amount of visual information (low, moderate, and high) as a within-subjects factor, and type of intention (non-social basic; non-social superordinate; social basic; social superordinate) and group (controls versus patients) as between-subjects factors.

Effect of the amount of visual information

We also calculated a score reflecting the influence of the variation in amount of visual information on each participant's performance. This score was obtained, for each participant, in each task, by subtracting the proportion of correct responses obtained in the high visual information condition from that obtained in the low visual information condition. This score was then entered in a $2 \times 4 \times 2$ repeated-measures ANOVA with bias (likely versus unlikely intentions) as a within-subjects factor, and type of intention and group as between-subjects factors.

Relationship to clinical symptoms

Finally, regression analyses were conducted to evaluate the influence of patients' cognitive performance on their clinical symptoms. In particular, we assessed whether an abnormal dependence on the bias and/or on visual information was predictive of the symptom severity on the different dimensions of schizophrenia measured (scale for the assessment of negative symptoms, scale for the assessment of positive symptoms and disorganization scores). For each clinical score, we conducted regression analyses using either the 'bias effect' score or the 'visual information effect' score as predictor variables. We used both these raw scores (simple linear regressions), or their transformed values (simple non-linear regressions with logarithmic, polynomial or exponential transformations). Models with the highest adjusted R^2 and a $P < 0.05$ are reported.

Results

For each session, two-tailed *t*-tests were performed between the two unlikely (or future unlikely) intentions on both reaction times and hits. As no significant differences appeared (all four tasks: all $P > 0.05$; see Supplementary Figs 1, 2A and B), performances for these two unlikely intentions were pooled for subsequent analyses.

Hits and reaction times

Baseline session

Overt blocks (containing very high and constant amount of visual information): in all four tasks, patients performed as successfully as controls when the amount of visual information was very high (mean correct responses > 94.5%, SD < 3.9; between-group comparisons: all $P > 0.05$), revealing that patients and controls were equally successful in integrating the probability distributions associated with each intention.

Covert blocks (containing varying amounts of visual information): the 2 (group) × 2 (intention: f-likely versus f-unlikely) × 3 (visual information) ANOVAs performed on both non-social (Basic and Superordinate) tasks revealed that patients performed the task as successfully as control participants [main effect of group, all *F*'s (1,48) < 1.25, all $P > 0.26$]. Furthermore, there were no significant differences in hits and reaction times between the 'future' likely intention (i.e. the one that participants will be biased towards in the subsequent bias session) and the 'future' unlikely intention, indicating that prior to biasing, there was no *a priori* bias towards one intention over another [main effect of intention (f-likely versus f-unlikely): all *F*'s (1,48) < 0.03, all $P > 0.84$; group × intention interaction effect, all *F*'s(1,48) < 0.12, all $P > 0.72$].

As the amount of visual information increased, intentions were discriminated both faster and more successfully [main effect of visual information: all *F*'s (2,96) > 251.1, all $P < 0.001$]. This improvement did not differ between patients and comparison participants [group × visual information interaction effect, all *F*'s (2,96) < 0.56, all $P > 0.57$]. The group × intention × visual information interaction was not significant, indicating that increasing the amount of visual information improved both groups' performance equally, and independently of the type ('future' likely versus 'future' unlikely) of intention [all *F*'s(2,96) < 0.61, all $P > 0.54$] (Supplementary Fig. 1A and B).

The 2 (group) × 2 (intention: f-likely versus f-unlikely) × 3 (visual information) ANOVAs performed on both social (Basic and Superordinate) tasks revealed that patients tended to be less successful than controls at recognizing intentions [main effect of group, all *F*'s (1,47–48) > 2.93, all $P < 0.084$)]. More specifically, in the social superordinate task, we found a significant interaction effect for hits between group and intention ('future' likely versus 'future' unlikely) factors, indicating that, prior to being biased, control participants displayed an early preference towards inferring a TFT compared with other strategies, which was not found for the patient group [group × intention interaction, $F(1,48) = 6.3$, $P = 0.014$; *post hoc* test comparing TFT versus other strategies in control participants, $P < 0.001$; *post hoc* test comparing controls versus patients on responding TFT, $P = 0.034$] (Supplementary Fig. 3). In the social basic task, controls also inferred a TFT response more frequently than schizophrenic patients (two-sample *t*-test, $t = -2.38$, $P = 0.028$), but the group × intention interaction effect did not reach significance [$F(1,47) = 2.7$, $P = 0.11$].

In social basic and social superordinate tasks, the performance of both groups increased with the amount of visual information [main effect of visual information: all *F*'s(2,94–96) > 198.64, all $P < 0.001$], but that increase was larger for patients than for controls [group × visual information, all *F*'s(2,94–96) > 3.17, all $P < 0.05$]. This increase was due to patients inferring TFT less frequently in the condition of a low amount of visual information, whilst inferring TFT as often as controls for medium and high amounts (*post hoc* test comparing percentage of hits between controls versus patients for low amount of visual information, $P < 0.05$; no significant differences found for the other amounts). The group × intention × visual information interaction effect, however, was not significant [all *F*'s (2,94–96) < 0.17, all $P > 0.21$] (Supplementary Fig. 2A and B).

In summary, in all four tasks intentions were recognized both faster and more successfully as the amount of visual information increased. In the non-social tasks, hits and reaction times for 'future' likely and unlikely intentions did not differ between groups, whereas in the social tasks, control participants exhibited an early preference for TFT strategies, prior to assignment of any probabilistic bias. This preference for inferring TFT over alternative strategies was not found in patients, which may account for their tendency to perform less successfully than control participants in social tasks, even when probabilities were not manipulated. Importantly, control participants tended to make more TFT responses as the amount of visual information decreased. This resulted in 'mechanically' reducing differences in the rate of likely responses between all three (low, medium and high) amounts of visual information. This effect was not observed in patients, due to their initial lack of preference for TFT.

Bias session

Overt blocks (very high and constant amount of visual information): in all four tasks both controls and patients performed the task successfully when the amount of visual information was very high (mean correct responses > 95%, SD < 3.1; between-group comparisons: all $P > 0.05$), indicating that patients and controls were equally successful in integrating the (biased) probability distribution associated with each intention.

Covert blocks (varying amounts of visual information): four distinct 2 (group) × 2 (bias: likely versus unlikely) × 3 (visual information) ANOVAs have been performed on each task (non-social basic, non-social superordinate, social basic and social superordinate). In all four tasks, participants were both more accurate and faster in recognizing the likely intention [i.e. the intention whose probability of occurrence was increased at the expense of the other competing ones; main effect of bias: all F's$(1,47–48) > 33.41$, all $P < 0.001$]. Similarly, performance increased with the amount of visual information in all tasks [main effect of visual information, all F's$(2,94–96) > 181.8$, all $P < 0.001$]. This effect was significantly modulated by the bias factor [visual information × bias interaction effect, all F's $(2,94–96) > 15.62$, all $P < 0.001$], with participants responding more frequently toward the likely intention as the amount of visual information progressively decreased, a finding consistent with previous results (Chambon *et al.*, 2011).

In the non-social basic task, patients performed the task as successfully as control participants [main effect of group $F(1,48) = 0.3$, $P = 0.85$], while in the non-social superordinate condition they exhibited significantly poorer performances than controls [main effect of group $F(1,48) = 9.17$, $P = 0.003$]. In the non-social superordinate task, the group × bias interaction was significant [$F(1,48) = 5.47$, $P = 0.023$]. Interestingly, decomposing this effect using *post hoc* Fisher tests revealed that patients chose the unlikely intention less frequently than controls ($P < 0.001$), but chose the likely intention as frequently as controls (Supplementary Fig. 4A). Furthermore, in this task, increasing the amount of visual information resulted in a larger increase of the rate of 'likely' responses for controls than for patients [group × visual information, $F(2,96) = 3.25$, $P = 0.04$; *post hoc* tests comparing per cent of hits between controls versus patients for medium and high amounts of information, all $P < 0.05$; no significant difference was found for the low amount]. No significant group × bias or group × visual information interactions were found in the non-social basic task.

In both social tasks, patients tended to recognize intentions less successfully than control participants [main effect of group all F's $(1,47–48) > 2.8$, $P < 0.09$]. The group × bias interaction effect was significant in the social superordinate task only [$F(1,48) = 6.37$, $P = 0.014$]: in this condition, patients were less likely to choose a TFT intention (least significant difference test, $P = 0.02$), while choosing the other, unlikely strategies as often as controls (Supplementary Fig. 4B). Crucially, we found that patients' performance increased to a larger extent than controls' as the amount of visual information increased in the two social tasks [group × visual information, all F's$(2,94–96) > 3.11$, all $P < 0.05$]. As in the baseline session, that increase was due to patients inferring TFT less frequently in the condition of a low amount of visual information (*post hoc* tests comparing per cent of hits between controls versus patients for the low amount of information, all $P < 0.05$; no significant differences were found for the other amounts). It is noteworthy that this pattern of performance (i.e. fewer responses toward TFT intentions and a greater effect of amount of visual information) was exactly the opposite of that observed in the non-social superordinate task.

Finally, we did not find any significant group × bias × visual information interaction effect in any of the four tasks [all F's $(2,94–96) < 2.32$, all $P > 0.1$], indicating that the controls' preference for TFT strategies was not modulated by the amount of visual information available (Supplementary Figs 1C and D, and 2C and D).

In summary, in all four tasks, both groups were more accurate and faster when responding toward the likely (i.e. biased) intention, and showed increased preference to this intention as the amount of visual information decreased. It is of note that this finding is consistent with predictions made by a Bayesian estimation scheme: in situations of sparse or incomplete data, participants tend to compensate for visual uncertainty by appealing to their prior knowledge (Chambon *et al.*, 2011).

In the non-social superordinate task, we found patients had difficulties in disengaging from their prior expectations (i.e. the likely intention) to select a response congruent with the unlikely intention, while relying less on visual information to make their decision. Such difficulties were associated with poor performances in this condition.

While patients also tended to perform less successfully than controls in the two social tasks, their performances however exhibited an opposite pattern of interaction between sensory and prior information: they were less sensitive than controls to the TFT bias, which resulted in their performance increasing to a larger extent than controls as a function of the amount of visual information.

Effect of the bias across different types of intentions

The 3 (visual information) × 4 (type of intention) × 2 (group) ANOVA first revealed a significant effect of the type of intention [main effect of type of intention, $F(3,191) = 11.7$, $P < 0.001$], with participants relying more on the bias to infer both social and non-social superordinate intentions, compared to basic ones (*post hoc* tests comparing superordinate and basic tasks, all $P < .001$). This difference interacted with group [group × intention interaction effect, $F(3,191) = 11.06$, $P < 0.001$]. Indeed, the bias exerted a greater influence on patients' response than controls' in the non-social superordinate condition ($P = 0.003$) while it exerted a smaller influence on their response compared to controls in both social conditions (all $P < 0.008$) (Fig. 3).

This difference could reflect the preference for TFT that controls already exhibited in the baseline session (see above), rather than reflecting a pure probabilistic bias effect. Therefore, to assess whether responses toward TFT strategies increased to the same extent with biasing across groups, we performed an additional 2 (group: controls, patients) × 2 (session: TFT-baseline, TFT-bias) repeated-measures ANOVA. No significant difference was observed [group × session interaction effect: both social conditions, all F's $(1,47–48) < 0.74$, all $P > 0.39$], indicating that the group difference for responding TFT in the bias session was due

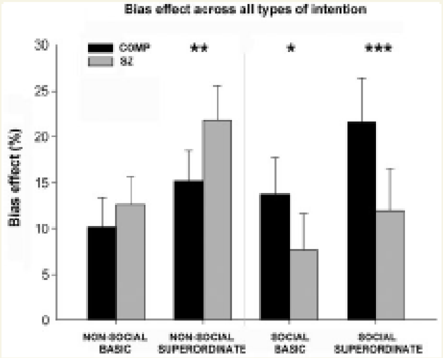


Figure 3 Bias effect (%) for all types of intention tasks. The greater the bias effect, the more participants respond toward the likely (i.e. biased) intention. SZ = patients with schizophrenia; COMP = comparison participants. *$P < 0.05$; **$P < 0.005$; ***$P < 0.001$.

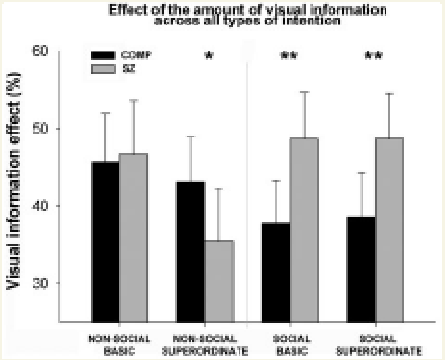


Figure 4 Effect of the amount of visual information (%) for all types of intention considered. The greater this score, the more participants' performance improved as the amount of visual information increased. SZ = patients with schizophrenia; COMP = comparison participants. *$P < 0.05$; **$P < 0.005$.

to controls' initial preference for responding TFT in the baseline session.

Effect of the amount of visual information across different types of intentions

The 2 (bias: likely versus unlikely) × 4 (type of intention) × 2 (group) ANOVA revealed a significant main effect of group [$F(1,191) = 4.12$, $P = 0.04$] showing that overall, patients' performances improved to a greater extent than controls' when increasing the amount of visual information. However, the significant interaction between group and type of intention [$F(3,191) = 6.12$, $P < 0.001$] further revealed that, while increasing the amount of visual information improved patients' performance more than controls' in both social basic and social superordinate conditions (*post hoc* tests, all $P < 0.005$), patients' performance improved to a lesser extent as this amount increased in the non-social superordinate condition (*post hoc* test, $P = 0.03$) (Fig. 4).

Clinical symptoms: regression analyses

Bias effect

In the non-social superordinate task, the bias effect significantly and positively predicted both scale for the assessment of positive symptoms ($R^2 = 0.39$, $P = 0.003$) and disorganization ($R^2 = 0.21$, $P = 0.04$) scores (Fig. 5A and B). The higher the effect of the bias on patients' performances (i.e. the more they relied on their prior knowledge to make their decision), the more likely patients were to be disorganized and exhibit positive symptoms. In both social tasks, the bias effect was found to significantly and negatively predict the scale for the assessment of negative symptoms score ($R^2 = 0.22$, $P = 0.03$ for the social basic task, and $R^2 = 0.32$, $P = 0.008$ for the social superordinate task). Therefore, the smaller the effect of the bias on patients' performances (i.e. the less they relied on their priors for inferring a social intention), the more severe the negative symptoms (Fig. 5C and D).

Effect of visual information

In the social superordinate intention task, the effect of visual information significantly predicted the scale for the assessment of negative symptoms score ($R^2 = 0.44$, $P = 0.001$) and—but to a lesser extent—the disorganization score ($R^2 = 0.28$, $P = 0.01$). The higher the effect of visual information on patients' responses (i.e. the more they relied on the visual information to make their decision), the more severe their negative and disorganization symptoms (Fig. 6B). In the social basic task, this effect tended to predict the scale for the assessment of negative symptoms score but the regression coefficient did not reach significance ($R^2 = 0.18$, $P = 0.066$) (Fig. 6A).

Discussion

The present study aimed to investigate whether the impaired ability of schizophrenic patients to appreciate other people's intentions is confined to a particular type of intention, as opposed to being generalized. To test this hypothesis, we designed a series of tasks that required the identification of different types of intentions, varying on the dimensions of scope (basic, superordinate) or target (non-social, social). We further hypothesized that, if present, a localized deficit would be accounted for by abnormalities in the interplay between prior knowledge and sensory evidence, which normally underlies the ability to infer others' intentions (Chambon *et al.*, 2011).

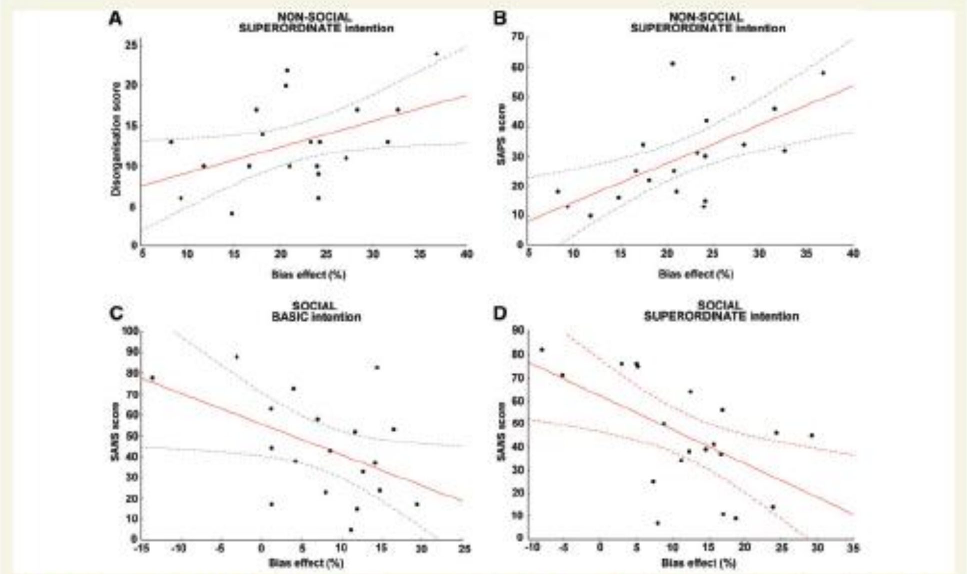


Figure 5 The magnitude of the bias effect predicts the severity of clinical symptoms. The linear regression lines derived from the linear regression analyses between the 'bias effect' (explanatory factor) and patients' disorganization (A), positive (B) and negative symptoms (C and D) are shown in red. Note that the correlation is positive for non-social intentions, and negative for social intentions. The 95% confidence intervals for the regression lines are shown in grey. SANS = Scale for the Assessment for the Negative Symptoms; SAPS = Scale for the Assessment of Positive Symptoms.

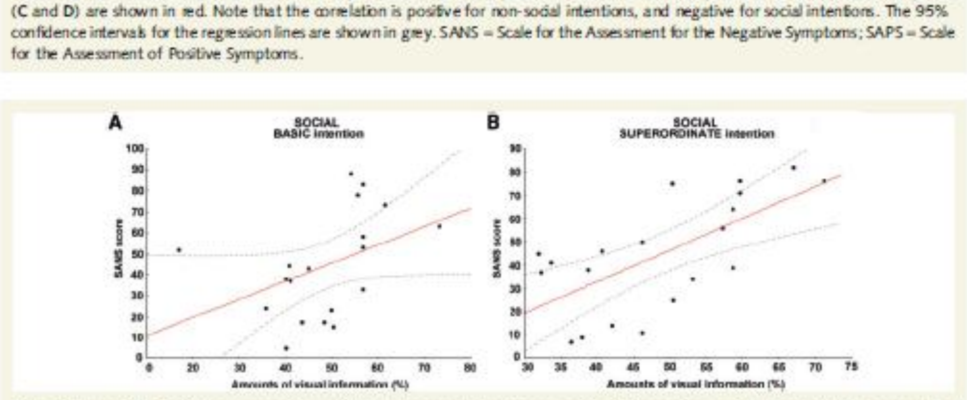


Figure 6 In social basic (A) and superordinate (B) tasks, abnormal dependence on visual information predicted scale for the assessment of negative symptoms. The linear regression lines derived from the linear regression analyses between the effect of the amount of visual information (explanatory factor) and patients' negative symptoms are shown in red. The 95% confidence intervals for the regression lines are shown in grey. SANS = Scale for the Assessment for the Negative Symptoms.

We first showed that controls and patients are sensitive to both types of information, and their interaction. First, all participants were more successful in recognizing underlying intentions when the visual information conveyed by the action scene was increased. Secondly, they showed more accuracy and were faster when recognizing likely compared to unlikely intentions. Finally, both groups' performances exhibited a strong bias effect, which progressively increased as the amount of visual information decreased, and vice versa. Crucially, we also observed specific differences between the two groups. Depending on the scope or target of the presented intention, patients with schizophrenia showed an abnormal weighting of either prior knowledge or sensory information, which was further associated with the severity of positive or negative symptoms of the condition.

Inferring non-social intentions

Whilst performing as successfully as controls when inferring intentions involving a single action (basic intentions), patients exhibited poorer performance when recognizing intentions involving a 'sequence' of basic motor acts (superordinate intentions). These poorer performances cannot be due to an increased attentional load, resulting from paying attention to a sequence of three actions as opposed to a single act, since patients were equally successful to controls in the baseline session of the superordinate condition. Indeed, differences between the patient and control groups were only found in the bias session, where the task required them to properly sample the probability distribution associated with each type of (likely or unlikely) intention.

Specifically, poorer performance in the bias session was characterized by a decreased number of responses toward unlikely intentions, with patients having difficulty disengaging from their biased expectations to select less likely alternatives. Crucially, this abnormal dependence on biased expectations may not be primarily due to a faulty weighting of probabilities—patients responded towards likely intentions as often as controls—but to an inability to revise prior expectations in light of new evidence. This assumption is supported by the fact that increasing the amount of visual information was of less benefit to patients' performances than controls', indicating that patients relied on visual information to a lesser extent than comparison participants to make their decision.

This inability to disengage from prior, self-generated expectations, together with a tendency to disregard external/sensory evidence, echoes specific biases observed in schizophrenia across a wide range of studies, such as the so-called 'bias against disconfirmatory evidence' (Woodward *et al.*, 2008), or a tendency to make hasty decisions ('jumping to conclusion': Garety *et al.*, 1991). Indeed, in tasks of probabilistic reasoning, individuals with schizophrenia tend to make judgements based upon less evidence than comparison participants and/or to hold prior beliefs despite little evidential support (Brankovic and Paunovic, 1999; Jones *et al.*, 1999). Our results in the superordinate condition reveal that this bias may not be specific to the domain of reasoning, but may also extend to mentalizing, potentially underlying such abnormalities in schizophrenia. When having to make decisions about other people's (non-social) intentions, patients preferentially relied on previously formed expectations, i.e. on beliefs about how the observed agent is most likely to behave, while neglecting potentially disconfirmatory visual information.

It is noteworthy that this pattern of performance was only observed in the superordinate, but not in the basic condition. We believe that this finding may be accounted for by the specific property of the intention manipulated. Indeed, a superordinate intention is achieved by a sequence of interchangeable basic actions, and, therefore, cannot be directly deduced from the current action *per se* (Pacherie, 2000; Jacob and Jeannerod, 2005). Inferring superordinate intentions thus requires the participant to refer to a distal representation of the goal, which is not directly available from observation and consequently tends to be less challenged by visual evidence (Chambon *et al.*, 2011). Intentions which are not predictable from merely observing the current action, i.e. superordinate intentions, may thus aggravate patients' tendency not to revise their beliefs in the face of progressively disconfirmatory evidence from the action scene.

An impaired revision process could signal a disturbance in the (Bayesian) inferential mechanism, which compares new sensory evidence with stored knowledge of the world, or prior beliefs (e.g. beliefs about what is the most likely cause of an observed behaviour, out of the possible alternatives). Under normal circumstances, a difference between expected and observed information gives rise to a prediction error that can be used to update one's model of the world (Kilner *et al.*, 2007*a*, *b*; Fletcher and Frith, 2009). A disturbance in this error-dependent updating mechanism, possibly caused by alterations in the dopaminergic circuitry (Gradin *et al.*, 2011), may result in patients having an abnormal degree of certainty in their beliefs about other people's intentions. If these beliefs are not challenged by external evidence and, if necessary, replaced with contextually appropriate beliefs, patients' inferences about others' mental states would be based on an outdated model of the current situation. This could result in patients exhibiting maladaptive or bizarre behaviours (Chambon *et al.*, 2008; Barbalat *et al.*, 2009) and/or holding incorrect (i.e. deluded) beliefs about the real causes driving other's behaviour (Fletcher and Frith, 2009). Precisely in line with this assumption, we found that patients' abnormal dependence on prior expectations predicted the severity of positive symptoms of schizophrenia: the more patients relied on their priors to make their decision, the more severe these symptoms were.

This observation sheds new light on previous evidence of patients with passivity symptoms misattributing actions to non-agents, or over-attributing intentionality where there is none (Abu-Akel and Bailey, 2000; Bentall *et al.*, 2001; Blakemore *et al.*, 2003), especially in situations that require continuous monitoring of visual signals arising from the action scene (Franck *et al.*, 2001). This 'hyper-intentionality' may indeed result from the quantitative over-generation of hypotheses and an inability to revise them in light of disconfirmatory evidence, potentially resulting in paranoid (over-)interpretation of other people's goals (Abu-Akel and Bailey, 2000; Bara *et al.*, 2011). Ultimately, over-reliance on unchallenged, internal expectations, whilst dismissing external evidence, could lead patients to make abnormal distinctions between the real causes driving other people's behaviour, and their subjective beliefs about what these causes should, or might be. Such a confusion between external and internal states of affairs would undermine a patient's ability to separate their own intentions from those of others (Frith and Corcoran, 1996), or to disentangle external experiences from inner experiences (Walter *et al.*, 2009), a feature that frequently accompanies paranoid and/or passivity symptoms, such as the well-documented 'delusion of control' (Brüne *et al.*, 2008).

Inferring social intentions

Surprisingly, the patients' pattern of performance on social tasks was the exact opposite of the pattern observed on non-social tasks. On social tasks, patients relied on the available visual information to a greater extent than controls, whilst showing a decreased sensitivity to the bias (i.e. towards TFT intentions). Furthermore, this pattern of performance predicted the severity of negative symptoms: patients with more severe negative symptoms were less sensitive to the bias (i.e. they were less likely to select TFT) and conversely, were more reliant on visual information to make their decision.

At first sight, a lower sensitivity to the bias in the social conditions may seem at odds with patients' excessive reliance on prior knowledge in the previous (non-social) conditions. However, performance in the bias session indicated that patients did normally integrate the (biased) probability distribution of the session, with the number of responses toward TFT increasing with its probability of occurrence, as found in controls. Rather, we found that unlike control participants, patients did not exhibit any early preference for TFT in the baseline session—that is, prior to being biased toward this particular mode of interaction. It is noteworthy that while increasing the probability of TFT was of benefit to both groups, this increase was not enough to compensate for patients' initial deficit.

This absence of an inherent preference for TFT suggests that for patients, social situations may not prompt the same expectations as those typically observed in healthy participants. Indeed, situations identified as involving social interactions are prone to trigger domain-specific expectations concerning the way agents are likely to behave in such situations (Castelli *et al*., 2000; Scholl and Tremoulet, 2000; Kourtis *et al*., 2010). Under normal circumstances, these modular, high-level expectations may contribute to priority being given to some intentional causes at the expense of other, competing causes (e.g. cooperation in response to previous cooperation, defection in response to defection; Chambon *et al*., 2011). Reliance on these domain-specific, prior expectations, which can be induced even by basic movements (such as the relative movements of geometrical figures; Heider and Simmel, 1944), may prove crucial in situations of sparse data, or when sensory evidence is too noisy to guarantee accurate inference-making (Baker *et al*., 2006). Poor performance in social conditions suggests that patients lack the prior expectations which usually bias social inferences, consistent with a previous suggestion that impoverished social knowledge, from which these expectations may be derived, constitutes an intrinsic feature of schizophrenia (Cutting and Murphy, 1990).

Concomitantly, impoverished expectations within the social domain may account for why patients were excessively over-reliant on visual information. As previously suggested, inferring another agent's intention requires the adaptive integration of new external evidence into prior beliefs about the agent's goals and attitudes (Baker *et al*., 2006; Fletcher and Frith, 2009), which is contingent upon the relative reliability of these two sources of information (Chambon *et al*., 2011). Results in the social conditions suggest a disturbance in this integrative mechanism that is exactly the opposite of what was observed in non-social conditions: impoverished expectations within the social domain, resulting in a reduced ability to draw reliable internal predictions, prompted patients to over-weight external evidence.

Crucially, this abnormal (over-)weighting of visual information correlated with the negative symptoms of schizophrenia. Previous observations have similarly shown that patients with negative symptoms, such as anhedonia or alexithymia, tend to excessively focus on directly observable, external information, rather than inner experience (Taylor, 1994). Our results further suggest that these incapacitating features may be accounted for by an impaired ability to make reliable predictions about other's behaviour, rendering patients slaves 'to every (external) influence' (Frith, 1994). It is noteworthy that over-reliance on external evidence can be particularly harmful in social situations, in which many possible intentions are potentially congruent with what is observed, so that it is impossible to infer the agent's intention from environmental cues only. In such situations, it has been shown that participants tend to compensate for sensory uncertainty by appealing to prior knowledge (Csibra and Gergely, 2007). Impoverished prior knowledge in the social domain may therefore result in an incapability to reduce the intrinsic uncertainty of social world. This may have important consequences on how patients suffering from negative symptoms perceive other people's social attitudes and behaviours. Indeed, a pervasive and constant uncertainty may render any observed, or experienced, social interactions fruitless, and could ultimately be responsible for social and motivational disorders that are characteristic of negative symptoms of schizophrenia (Fletcher and Frith, 2009).

It is noteworthy that a lack of preference for TFT in schizophrenia is consistent with previous data. In tasks simulating human cooperation in group interactions, patients do not exhibit any pattern of 'equivalent retaliation', or 'altruistic punishment', e.g. they do not defect when the game partner has previously defected, or they accept unfair offers at a significantly higher rate than did healthy controls (Chung *et al*., 2011; Csukly *et al*., 2011). Interestingly, individuals with schizotypal traits exhibit the same pattern of performance as patients (van't Wout and Sanfey, 2011), suggesting that poor expectations in the social domain may represent a marker of vulnerability to schizophrenia. Moreover, such abnormal expectations may serve as early clinical intervention targets. Indeed, there is growing evidence that cognitive therapies targeting social skills improves long-term prognosis and significantly benefits the patient's everyday life (Horan *et al*., 2011; Ventura *et al*., 2011). We believe accordingly that early detection of an abnormal use of social-specific knowledge may have a positive impact on both patients' social functioning and evolution of the condition. Critically, a lack of preference for the 'TFT' mode of interaction predicted the severity of negative but not positive symptoms. This further suggests that poor expectations in the social domain may also be relevant to the formation of symptom profile, together with being a useful indicator for identification and management of vulnerable individuals.

Neural underpinnings

Together, these results suggest that different neural dysfunctions may underline patients' abnormal performance, depending on their symptom profile and the type of intention considered. According to a recent model, action understanding is achieved through interactions between a ventral pathway where intention priors are formed, and a dorsal network hierarchically organized according to the level at which the observed action is represented (kinematic, motor command, or goal level) (Kilner, 2011). Intention priors in the ventral pathway are used to predict in dorsal areas the most likely action required to achieve the most likely intention, given what is observed. An error signal is generated when the prediction is not accurate. We suggest that undue weight given to prior expectations in patients with positive symptoms may be caused by abnormal encoding of prediction error signals in dopamine-rich brain areas (Gradin *et al.*, 2011). This would result in the inability to update intention priors in brain areas of the ventral pathway. On the other hand, lack of preference for TFT in patients with negative symptoms suggests abnormal biasing influences from brain regions that encode social-specific knowledge, such as the medial prefrontal cortex (Overwalle, 2009). Thus, the medial prefrontal cortex might insufficiently bias action prediction in brain areas within the dorsal pathway, resulting in an equal weighting of all possible action alternatives (e.g. cooperation if previous defection, cooperation if previous cooperation, etc.). Such weakening of social-specific influences is likely to reduce the accuracy of prediction error-dependent mechanisms, leading patients to rely on sensory evidence by default. In future work, the use of neuroimaging techniques should allow us to test these assumptions directly.

Conclusion

We identified specific mentalizing impairments in participants with schizophrenia. Rather than being generalized, these impairments were contingent upon the scope (basic versus superordinate) or the target (non-social versus social) of the intention to be inferred, and were further accounted for by abnormal integration of visual information and prior knowledge.

In non-social tasks, patients showed specific difficulties in inferring intentions achieved by a sequence of basic motor acts (superordinate intentions). We found that this poor performance was due to patients over-relying on prior expectations and disconfirming visual evidence. This abnormal pattern of interaction predicted the severity of positive symptoms. We suggested that this faulty interaction may signal a disturbance in the inferential mechanism driving the integration of sensory evidence into prior beliefs, to produce accurate inference about other people's intentions. Such a disturbance could favour a paranoid (over-) interpretation of other people's goals, by hindering the revision of one's prior beliefs, and may ultimately lead patients to distinguish abnormally between their own and others' intentions—a confusion frequently experienced by individuals with passivity symptoms.

Patients also showed difficulties in inferring social intentions. However, their pattern of performance was the exact opposite to that observed in non-social conditions. While they exhibited weaker prior expectations, they relied strongly on sensory evidence to make their decisions. Furthermore, this pattern of performance predicted the severity of negative symptoms. Based on the absence of early preference for the TFT mode of interaction, we hypothesized that social situations may not prompt the same expectations in patients as those typically observed in healthy participants, leading to the formation of abnormal (unreliable) predictions about others' social intentions. Such abnormal predictions may result in an incapability to reduce the intrinsic uncertainty of social situations. We suggest that constant and pervasive uncertainty about other's social attitudes and behaviours could jeopardize patients' propensity to social interactions, and may ultimately account for some of the incapacitating features associated with negative symptoms of schizophrenia.

Acknowledgements

We are grateful to Mathilde Vernet, Juliette Sablier and Mariia Kaliuzhna for their help in preparing the stimuli. We also gratefully acknowledge Mrs Ghislaine Bailly, Drs Olivier Martin and Caroline Demily and Pr. Jean-Louis Terra for their help in the inclusion of patients. We thank Emma Kilford and Geeta Reddy for their help in proofreading the final draft of the article.

Funding

Agence Nationale de la Recherche obtained by C.F. (ANR-ZR81); Conseil Scientifique de Recherche, Le Vinatier (CSR U08); French Ministry of Research (to V.C.).

Supplementary material

Supplementary material is available at *Brain* online.

References

Abu-Akel A, Bailey AL. The possibility of different forms of theory of mind. Psychol Med 2000; 30: 735–8.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, DSM-IV. 4th edn. Washington, DC: APA; 1994.

André JB, Day T. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. J Theor Biol 2007; 247: 11–22.

Andreasen NC. The scale for the assessment of negative symptoms (SANS). Iowa City: The University of Iowa; 1983.

Andreasen NC. The scale for the assessment of positive symptoms (SAPS). Iowa City: The University of Iowa; 1984.

Axelrod R. The complexity of cooperation. Princeton, NJ: Princeton University Press; 1997.

Baker CL, Tenenbaum JB, Saxe RR. Bayesian models of human action understanding. In: Weiss Y, Scholkopf B, Platt J, editors. Advances in neural information processing systems. Cambridge, MA: MIT Press; 2006. p. 99–106.

Baker CL, Saxe RR, Tenenbaum JB. Action understanding as inverse planning. Cognition 2009; 113: 329–49.

Bara BG, Ciaramidaro A, Walter H, Adenzato M. Intentional minds: a philosophical analysis of intention tested through fMRI experiments involving people with schizophrenia, people with autism, and healthy individuals. Front Hum Neurosci 2011; 5: 7.

Barbalat G, Chambon V, Franck N, Koechlin E, Farrer C. Organization of cognitive control within lateral prefrontal cortex in schizophrenia. Arch Gen Psych 2009; 66: 1–10.

Bentall RP, Corcoran R, Howard R, Blackwood N, Kinderman P. Persecutory delusions: a review and theoretical integration. Clin Psychol Rev 2001; 21: 1143–92.

Blakemore SJ, Frith U. How does the brain deal with the social world? Neuroreport 2004; 19: 119–28.

Blakemore SJ, Boyer P, Pachot-Clouard M, Meltzoff A, Segebarth C, Decety J. The detection of contingency and animacy from simple animations in the human brain. Cereb cortex 2003; 13: 837–44.

Bora E, Yucel M, Pantelis C. Theory of mind impairment in schizophrenia: meta-analysis. Schizophr Res 2009; 109: 1–9.

Brankovic SB, Paunovic VR. Reasoning under uncertainty in deluded schizophrenic patients: a longitudinal study. Eur Psychiatry 1999; 14: 76–83.

Brüne M, Lissek S, Fuchs N, Witthaus H, Peters S, Nicolas V, et al. An fMRI study of theory of mind in schizophrenia patients with 'passivity' symptoms. Neuropsychologia 2008; 46: 1992–2001.

Castelli F, Happé F, Frith U, Frith CD. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 2000; 12: 314–25.

Chambon V, Domenech P, Pacherie E, Koechlin E, Baraduc P, Farrer C. What are they up to? The role of sensory evidence and prior knowledge in action understanding. Plos ONE 2011; 6: e17133.

Chambon V, Franck N, Koechlin E, Ciuperca G, Fakra E, Azorin J-M, et al. The architecture of cognitive control in schizophrenia. Brain 2008; 131: 962–70.

Chung D, Kim YT, Jeong J. Cognitive motivations of free riding and cooperation and impaired strategic decision making in schizophrenia during a public goods game. Schizophr Bull 2011; 24.

Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, et al. The intentional network: how the brain reads varieties of intentions. Neuropsychologia 2007; 45: 3105–13.

Corcoran R. Theory of mind in schizophrenia. In: Penn D, Corrigan P, editors. Social cognition in schizophrenia. Washington: APA; 2001.

Csibra G, Gergely G. Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol 2007; 124: 60–78.

Csukly G, Polgár P, Tombor L, Réthelyi J, Kéri S. Are patients with schizophrenia rational maximizers? Evidence from an ultimatum game study. Psychiatry Res 2011; 187: 11–7.

Cutting J, Murphy D. Impaired ability of schizophrenic, relative to manics or depressives, to appreciate social knowledge about their culture. Br. J. Psychiatry 1990; 157: 355–8.

Franck N, Farrer C, Georgieff N, Marie-Cardine M, Daléry J, d'Amato T, et al. Defective recognition of one's own actions in patients with schizophrenia. Am J Psychiatry 2001; 158: 454–9.

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10: 48–58.

Frith CD. Theory of mind in Schizophrenia. In: David AS, Cutting JC, editors. The neuropsychology of schizophrenia. Hillsdale, NJ: Erlbaum Press; 1994. p. 147–61.

Frith CD, Corcoran R. Exploring 'theory of mind' in people with schizophrenia. Psychol Med 1996; 26: 521–30.

Frith CD. Schizophrenia and theory of mind. Psychol Med 2004; 34: 385–9.

Garety PA, Hemsley DR, Wessely S. Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. J Nerv Ment Dis 1991; 179: 194–201.

Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, et al. Expected value and prediction error abnormalities in depression and schizophrenia. Brain 2011; 134: 1751–64.

Griffiths TL, Kemp C, Tenenbaum JB. Bayesian models of cognition. In: Sun R, editor. The Cambridge handbook of computational cognitive modelling. Cambridge: Cambridge University Press; 2008.

Hardy-Baylé MC, Sarfati Y, Passerieux C. The cognitive basis of disorganization symptomatology in schizophrenia and its clinical correlates: toward a pathogenetic approach to disorganization. Schizophr Bull 2003; 29: 459–71.

Harrington L, Siegert RJ, McClure J. Theory of mind in schizophrenia: a critical review. Cognit Neuropsychiatry 2005; 10: 249–86.

Heider F, Simmel M. An experimental study of apparent behaviour. Am J Psychol 1944; 57: 243–59.

Hemsley DR. The development of a cognitive model of schizophrenia: placing it in context. Neurosci Biobehav Rev 2005; 29: 977–88.

Horan WP, Green MF, Degroot M, Fiske A, Hellemann G, Kee K, et al. Social cognition in schizophrenia, Part 2: 12-month stability and prediction of functional outcome in first-episode patients. Schizophr Bull 2011 (in press).

Jacob P, Jeannerod M. The motor theory of social cognition: a critique. Trends Cogn Sci 2005; 9: 21–5.

Jones E. The phenomenology of abnormal belief: a philosophical and psychiatric inquiry. Phil Psych Psychol 1999; 6: 1–16.

Kourtis D, Sebanz N, Knoblich G. Favouritism in the motor system: social interaction modulates action simulation. Biol Lett 2010; 6: 758–61.

Kilner JM. More than one pathway to action understanding. Trends Cogn Sci 2011; 15: 352–7.

Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. Cogn Process 2007a; 8: 159–66.

Kilner JM, Friston KJ, Frith CD. The mirror-neuron system: a Bayesian perspective. Neuroreport 2007b; 18: 619–23.

McCabe R. On the inadequacies of theory of mind explanations of schizophrenia. Alternative accounts of alternative problems. Theory Psychol 2004; 14: 738–52.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971; 9: 97–113.

Pacherie E. The content of intentions. Mind Language 2000; 15: 400–32.

Pacherie E. The phenomenology of action: a conceptual framework. Cognition 2008; 107: 179–217.

Scholl BJ, Tremoulet PD. Perceptual causality and animacy. Trends Cogn Sci 2000; 4: 299–309.

Sprong M, Schothorst P, Vost E, Hox J, van Engeland H. Theory of mind in schizophrenia: a meta-analysis. Br J Psychiatry 2007; 191: 5–13.

Taylor GJ. The alexithymia construct: conceptualization, validation, and relationship with basic dimensions of personality. New Trends Exp Clin Psychiatry 1994; 10: 61–74.

van Overwalle F. Social cognition and the brain: a meta-analysis. Hum Brain Mapp 2009; 30: 829–58.

van't Wout M, Sanfey AG. Interactive decision-making in people with schizotypal traits: a game theory approach. Psychiatry Res 2011; 185: 92–6.

Ventura J, Wood RC, Hellemann GS. Symptom domains and neurocognitive functioning can help differentiate social cognitive processes in schizophrenia: a meta-analysis. Schizophr Bull 2011 (in press).

Walter H, Ciaramidaro A, Adenzato M, Vasic N, Ardito RB, Erk S, et al. Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI type. Scan 2009; 4: 166–76.

Woodward T S, Moritz S, Menon M, Klinge R. Belief inflexibility in schizophrenia. Cognit. Neuropsychiatry 2008; 13: 267–77.

Zalla T, Bouchilloux N, Labruyere N, Georgieff N, Bougerol T, Franck N. Impairment in event sequencing in disorganised and non-disorganised patients with schizophrenia. Br Res Bull 2006; 68: 195–202.

doi:10.1093/brain/aws169 Brain 2012: Page 1 of 2 | e1

BRAIN
A JOURNAL OF NEUROLOGY

LETTER TO THE EDITOR

Reply: The Bayesian equation and psychosis

Valerian Chambon,[1] Philippe D. Domenech,[2] Guillaume Barbalat,[3] Elisabeth Pacherie,[4]
Pierre Jacquet[5] and Chlöé Farrer[1]

1 Centre de Neuroscience Cognitive, 67 bd Pinel, 69675 Bron, France
2 CNRS—Cognitive Neuroscience Centre, Lyon, France
3 Laboratoire Langage, Cerveau, Cognition, Bron, France
4 Institut Jean Nicod—DEC-ENS, Paris, France
5 Department of Psychology, University of Bologna, Bologna, Italy

Correspondence to: Valerian Chambon,
Centre de Neuroscience Cognitive,
67 bd Pinel, 69675 Bron, France
E-mail: valerian.chambon@isc.cnrs.fr

Sir, In a recent study (Chambon *et al*., 2011), we investigated the ability of patients with schizophrenia to make accurate predictions about other people's intentions. This ability has long been shown to be impaired in schizophrenia, and this impairment may be accounted for by an abnormal integration of two different sources of information: the sensory evidence conveyed by movement kinematics, and the observer's expectations about how likely an intention is. In the task, these two types of information were systematically varied. Our results showed that patients with positive symptoms were prone to over-weight sensory evidence confirming their prior expectations and to disregard evidence that invalidated such priors. We hypothesized that this abnormal interplay of prior expectations and current sensory experiences—that normally guarantee accurate inferences—could result from a disturbance in prediction error signalling, possibly caused by alterations in the dopaminergic circuitry. We speculated that such aberrant prediction error signals might account for the formation, as well as for the update, of delusional beliefs as to how biological agents are most likely to behave.

In their comment, Garrett and Singh suggest that our results cannot be readily applied *in vivo* to psychosis 'without distinguishing between pre-psychotic and post-psychotic prior experience'. Specifically, they argue that belief updating should be differentially impaired according to whether sensory evidence combines with a belief that has been formed prior to, or during, a psychotic episode. They argue that, in clinical psychosis, patients do not abnormally under- or over-weight information indistinctly, but that evidence disconfirming 'psychotic' beliefs only is under-weighted, whereas evidence confirming these beliefs is given too much credit. During acute psychotic episodes, confirmatory evidence is over-weighted because such evidence (e.g. 'the fleeting glance of a stranger') readily combines with previous confirmatory observations (e.g. 'strangers are wearing blue coats') to progressively inflate the veracity of the initial delusional belief (i.e. 'FBI agents are watching me').

This comment raises two important points. First, it suggests that the integration of sensory evidence and prior beliefs somehow differs according to whether these beliefs are 'psychotic' in essence, i.e. whether they involve a delusional content, or not. Second, the formation of delusional beliefs results from a deviation of this integration process relative to the 'Bayesian norm'. In Garrett and Singh's view, this process of delusion formation qualitatively differs from the process whereby formed beliefs (both delusional and non-delusional) are maintained during acute psychotic episode. As our task did not contrast delusional versus non-delusional beliefs, Garrett and Singh argue that our results pinpoint a disturbance in a non-specific mechanism of belief maintenance, but say little about how ordinary beliefs progressively convert into delusional thoughts.

We fully agree with Garrett and Singh that the formation of delusional beliefs calls for specific investigations, as it is still unclear whether it entails specific computational processes. For example, delusional beliefs might arise as a consequence of either misassessed prior probabilities (e.g. delusion-related priors may be abnormally inflated in patients), or misaggregated components of the Bayes' rule (i.e. priors and likelihood may not combine according to this rule), or both (McKay, 2012). However, the fact that such processes specifically account for the emergence of delusional beliefs does not preclude their participation in the abnormal maintenance of beliefs. Indeed, both delusion formation and abnormal persistence of beliefs may just be explained by a single neurophysiopathological mechanism. Aberrant prediction error signalling in the context of dopamine dysregulation has been proposed as a probable candidate (Corlett *et al*., 2009; Fletcher and Frith, 2009). In a nutshell, prediction error is defined as the discrepancy between what is expected and what is actually experienced. Such errors signal the need to update one's beliefs about the world. Under physiological conditions, individuals update beliefs in an approximately Bayesian fashion, whereas pathological prediction error signals may yield significant deviations from normative Bayesian inference (Friston, 2010; McKay, 2012). It has been suggested that aberrant coding of prediction error in deluded patients may abnormally over-weight random events. Delusions would thus arise as a means of explaining, or accommodating, these odd, and abnormally salient experiences (Kapur, 2003; Corlett *et al*., 2009). Interestingly, the persistence of delusion, as well as the abnormal inflexibility of otherwise non-delusional beliefs, can be explained through the same mechanism. Indeed, in Bayesian hierarchical networks, persistent delusional beliefs are more likely as one goes up towards more abstract hierarchical levels, which depend less on external inputs and more on previous stage representations (Friston, 2010). Thus, misbeliefs can persist because of persistent abnormal prediction error throughout the whole hierarchy. A belief induced by the coincidental hypersalience of an event would thus be progressively reinforced and enriched as aberrant prediction error randomly occurs (Corlett *et al*., 2009). When the psychotic equilibrium is reached, persistent error signals would continuously reconsolidate this initial explanation, up to a point where the belief in this explanation becomes absolutely impervious to contradictory evidence, and ultimately creates the framework within which all perceptual experiences are interpreted.

In sum, we agree with Garrett and Singh that delusional beliefs may arise as a consequence of multiple deviations from normative Bayesian updating. However, we do not think there is a need to postulate a separate mechanism to explain why, at the same time, non-delusional (pre-psychotic) beliefs may prove highly inflexible and abnormally persist in the face of contradictory evidence.

Acknowledgements

We wish to thank Prof. Paul Fletcher for useful discussions on some of the issues addressed in this letter.

References

Chambon V, Pacherie E, Barbalat G, Jacquet P, Franck N, Farrer C. Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia. Brain 2011; 134: 3725–38.

Corlett PR, Krystal JH, Taylor JR, Fletcher PC. Why do delusions persist? Front Hum Neurosci 2009; 3: 1–9.

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10: 48–58.

Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010; 11: 127–38.

Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 2003; 160: 13–23.

McKay R. Delusional inference. Mind Lang 2012. (in press).

Abstract and Concrete Sentences Processing Differentially Modulates Corticospinal Excitability

(under revision in Brain Research)

Claudia Scorolli[1], Pierre Olivier Jacquet[1], Ferdinand Binkofski[2],
Roberto Nicoletti[3], Alessia Tessari[1], & Anna M. Borghi[1, 4]

[1]Department of Psychology, University of Bologna, [2] Division for Cognitive Neurology, RWTH Aachen, [3]Department of Communication Disciplines, University of Bologna,
[4]Institute of Cognitive Sciences and Technologies, CNR, Rome

Corresponding author:
Claudia Scorolli
Department of Psychology, Viale Berti Pichat, 5, 40100, Bologna, Italy; Voice: +39-051-2091838;
e-mail: claudia.scorolli2@unibo.it

Abstract: 250 words.
Article length: 33 pages, 8763 words, including references.
Figures: 4.
Running title: A TMS study on Abstract and Concrete Sentences

Abstract

An important challenge of embodied theories is to explain the comprehension of abstract sentences. The aim of the present study was to scrutinize the role of the motor cortex in this process. We developed a new paradigm to study the abstract-concrete dimension by combining action-related and non-action-related verbs with nouns of graspable and non-graspable objects. Using these verb-noun combinations we performed a Transcranial Magnetic Stimulation (TMS) on the left primary motor cortex while participants performed a sentence sensibility task. Single-TMS pulses were delivered 250 ms after verb or noun presentation in each of four combinations of Abstract and Concrete verbs and nouns. To evaluate corticospinal excitability we registered the electromyographic activity of the right first dorsal interosseous muscle. As to verb-noun integration, analysis of motor evoked potentials (MEPs) after TMS pulse during noun presentation revealed greater peak-to-peak amplitude in sentences containing Abstract rather than Concrete Verbs. Response times were also collected and showed that compatible (Concrete-Concrete and Abstract-Abstract) combinations were processed faster than mixed ones; moreover in combinations containing concrete verbs, participants were faster when the pulse was delivered on the first word (verb) than on the second one (noun). Results support previous findings showing precocious activation of hand-related areas after concrete verbs processing. The prolonged or delayed activation of the same areas by abstract verbs will be discussed in the framework of recent embodied theories based on multiple types of representation, particularly theories emphasizing the role of different acquisition mechanisms for concrete and abstract words (Borghi & Cimatti, 2009;2012).

Keywords: embodied cognition, primary motor cortex, language grounding, abstract and concrete sentences.

1. INTRODUCTION

The ability to understand and use abstract words is an important part of the human capacity to interact with the environment and with others. While many studies have been devoted to this important topic, the issue of how abstract concepts and words are represented is still unsolved (for a recent review, see Pecher et al., 2011). It is well known that abstract words are remembered and recognized more slowly than concrete ones (Schwanenflugel, 1991). Their processing can engage mental imagery, but at a lower rate and with a greater variability compared to concrete words (Paivio, et al., 1968; Paivio, 1991). It is also well established that abstract words are acquired later than concrete and generally highly imaginable words (Bird et al., 2001). Finally, the double dissociations found between the understanding of abstract and concrete words (Shallice & Warrington, 1975; Warrington, 1975) further suggest that, even if the domain of ‘abstract concepts’ is not homogeneous, there must be some common features that link its variegated members.

In recent years many neuroimaging and meta-analyses have investigated the differing neural correlates involved in abstract and concrete concepts (for a recent quantitative meta-analysis see Wang et al., 2010). On one hand this concern is due to a genuine interest in the specific topic, on the other hand this interest is strongly related to the theoretical implications of this issue for embodied and grounded theories of cognition (for a review on different kinds of embodied views, see Goldman & De Vignemont, 2009). Embodied theories vary in their details, but most of them maintain that all concepts and words activate a simulation mechanism that recruits the same action, perception and emotional networks activated during actual experience with their referents (e.g. Barsalou, 1999; 2003; Glenberg & Robertson, 2000; Zwaan, 2004). Notice that different versions of the notion of simulation have been proposed (for reviews, see Borghi, in press; Decety & Grezes, 2006). The term “simulation” as we intend it here involves two aspects: it implies the re-enactment of past experiences (Barsalou, 1999) and it is predictive. It refers to a process that is embodied, unconscious, not deliberate, and it is aimed at action preparation (Gallese, 2009). In contrast with other views (e.g., Decety & Ingvar, 1990) simulating does not imply a deliberate reactivation of previously performed actions, and it does not consist in a posteriori forms of motor imagery. While empirical evidence on simulation is compelling with respect to concrete concepts and words, the challenge these theories face with is to clarify whether abstract concepts and words are also represented via embodied simulations. Mental metaphors could represent a potential solution, as they import the image-schemas derived from the source domain of sensorimotor experience (Lakoff, 1987; Gibbs & Steen, 1999). Compelling evidence has been collected in favor of this approach (e.g., Casasanto, 2009), but it is hard to foresee how it can be generalized to all varieties of abstract words.

Recently, some scholars have addressed the issue by getting to the root of the problem: embodied accounts of language have focused largely on language grounded in bodily experiences but have neglected that language also plays a role in shaping our experience (Borghi & Cimatti, 2009; 2012; Borghi & Pecher, 2011). In their proposal (Words as Tools, WAT) Borghi and Cimatti (2009; 2012), similarly to other authors (Dove, 2009, 2010; Louwerse & Jeauniaux, 2009; Barsalou et al., 2008; Simmons et al., 2008), try to integrate linguistic and modal approaches. The unique quality of the WAT proposal maintains that the linguistic system does not simply involve a form of superficial processing and that words cannot be conceived of as mere signals of something. Words are also tools that allow us to operate in the world (Clark, 2007; Mirolli & Parisi, 2011; Tylèn et al., 2010). The WAT proposal has direct implications for the explanation of abstract word meanings. Indeed, Borghi and Cimatti (2009; 2012) proposed that, probably due to their different acquisition mechanisms, abstract word meanings rely on the social experience of language. With concrete words, such as "phone", the word's referent can be indicated and tagged using linguistic labels. With abstract words, instead, there is not a specific referent to be indicated. In this case, the word, such as "freedom", plays a major role, as it helps assemble a set of diverse sensorimotor experiences (e.g., we put together different experiences of freedom once we have learned the word "freedom"). In support of this proposal, Borghi et al. (2011) have shown that the acquisition modality of novel concrete and abstract words (manipulation of their referents vs. simply visualization of scenes with interacting objects) has an impact on their representation: in a verification task participants responded faster to abstract words when using the microphone, and to concrete words when using the keyboard. The results indicate that concrete words evoke more manual information, whereas abstract words evoke more linguistic information; importantly, the advantage of the microphone with abstract words was more pronounced when the meaning of the word was linguistically explained, and it was not present when the linguistic information contrasted the perceptual information. These results clearly show the similarities but also the differences between embodied accounts (Barsalou et al., 2008; Borghi & Cimatti, 2009; 2012; Simmons et al., 2008; for recent brain imaging evidence consistent with this view see Rodríguez-Ferreiro, et al., 2010) and Paivio's dual coding theory (e.g, Paivio, 1986; Binder et al., 2005; Desai et al., 2010). Both accounts share the idea that multiple types of representation underlie knowledge, but embodied proposals differ from Paivio's view as they hypothesize that not only concrete, but also abstract words are grounded in perception and action.

The aim of the present study is to test the WAT proposal (Borghi & Cimatti, 2009; 2010; 2012) through scrutinizing the possible modulation of the left primary motor cortex (M1) activity during abstract and concrete sentence processing. We used an innovative paradigm recently developed by Scorolli et al. (2011), in which the same Concrete Verb (CV) was combined with a Concrete Noun (CN) and with an Abstract Noun (AN), the same Abstract Verb (AV) was combined with the nouns previously used. One of the advantages of this design is the possibility to study abstractness along a continuum - that is, to study combinations in which abstract and concrete verbs and nouns are put together. This paradigm was adapted to the use of single-pulse transcranial magnetic stimulation (TMS) technique, with the aim to explore the modulation of M1 activity during the processing of action-related and non-action-related verbs, combined with nouns of graspable and non-graspable objects.

Resting on the predictions of the WAT proposal, we hypothesized that the processing of language is different within the motor cortex for concrete and abstract language content. On the basis of the assumption that the mode and age of acquisition of concrete and abstract words differ, we expected to also find clues for different roots of processing. Specifically, our predictions are:

I. Given that according to embodied theories both concrete and abstract words are grounded in the motor system, we predict that concrete and abstract words differentially recruit neurons of the hand areas in M1 (detectable on the modulation of motor evoked potentials, MEPs, analyses).

II. If concrete words, and concrete verbs in particular, evoke motor information more directly than abstract words, we predict:

IIa. an earlier activation of hand representation areas in concrete verb processing over abstract verb processing, detectable on MEPs collected after a pulse delivered on the first word, and a later modulation due to abstract verb processing, detectable on MEPs collected after a pulse delivered on the second word;

IIb. faster sentence processing when the pulse is delivered on concrete verbs than on abstract verbs (detectable on response times, RTs, analyses).

These effects should be present only when the sentences are sensible, otherwise no simulation should occur, or the simulation should be interrupted when the first word (verb) has to be combined with the second one (the noun).

2. RESULTS

Our dependent variables were reaction times (RTs) and motor evoked potentials (MEPs). It is worth noting that, due to the fact that we used sentences instead of single words and that for each sentence we stimulated either the verb or the noun, results from these two measurements cannot be completely matched. To clarify: I. MEPs recorded after the stimulation on the first word (verb) provide information concerning the first part of sentence processing (the processing of a verb that has later to be integrated with a specific noun); II. MEPs recorded after stimulation on the second word (noun) provide us with information on the integration between the verb and the noun (whole sentence processing) ; III. RTs provide information on the whole sentence processing.

3.1 Analyses on MEPs

One participant was eliminated from analyses as, due to reported high levels of anxiety, we stopped the experimental session before finishing the overall experiment. As predicted in the Sham condition we did not record any MEPs, so we will not further discuss the non-active condition. Peak-to-peak amplitude (mV) of each MEP was computed by an automatic Excel script prior to normalization by means of a logarithmic transformation [log10 (mean MEPs amplitude value)]. MEP amplitudes inferior to 0.05 mV were excluded from analyses. One participant was excluded from further analyses due to the high percentage of unrecorded MEPs (25.45 %). We eliminated MEPs for which participants gave an incorrect response on the sentence sensibility task.

Normalized MEPs recorded after TMS stimulation on the first word (verb) were submitted to a t-test, with Verb (Concrete vs. Abstract) working as the within participant variable. MEPs peak-to-peak amplitudes recorded from the right FDI muscle during TMS delivery did not differ in the case of Concrete Verbs or Abstract Verbs ($p = 0.19$).

Normalized MEPs recorded from the right FDI muscle after the stimulation on the second word (noun) provided information on the verb and noun integration. This allowed for a 2 (Verb: Concrete vs. Abstract) X 2 (Noun: Concrete vs. Abstract) ANOVA, with all variables manipulated within participants. We eliminated MEPs for which participants gave an incorrect response on the sentence sensibility task. We found a significant main effect of the Verb, $F(1, 13) = 13.21$, $MSe = 0.002$, $p < .005$: in case of active pulse, peak-to-peak MEPs amplitude was greater for sentences containing Abstract Verbs (M = 2.71) than for sentences containing Concrete Verbs (M = 2.67, see Fig. 1).

The last result obtained when the pulse was delivered on 2nd word (the noun) shows that the primary motor cortex activity is specifically modulated by the processing of Abstract Verbs. Overall, this result gives an additional hint as to the recruitment of the motor system during Abstract Verbs processing.. To understand if this recruitment occurs later or lasts longer than with Concrete Verbs, we contrasted the kind of verb and the timings of TMS delivery. As we found no effect of the kind of noun, we were entitled to perform a 2 (Stimulated Word: 1st word vs. 2nd word) X 2 (Verb: Concrete vs. Abstract) ANOVA. We found a significant interaction between the Pulse and the Verb, F (1, 27) = 13.78, MSe = 0.001, $p < .001$: abstract verbs obtained greater peak-to-peak MEPs amplitude when the pulse was delivered 650 ms (400+250), M = 2.71, rather than 250 ms, M = 2.67, after the verb presentation, post hoc LSD: $p < .005$. Symmetrically we found that concrete verbs obtained greater peak-to-peak MEPs amplitude for the first timing of TMS delivery, M = 2.69, than for the second one (650 ms), M = 2.67, post hoc LSD: p = .055. Interestingly the activation of the motor system for concrete verb after a precocious pulse did not differ from the one obtained for abstract verb after a delayed pulse (M = 2.69 vs. M = 2.71, p = .07).

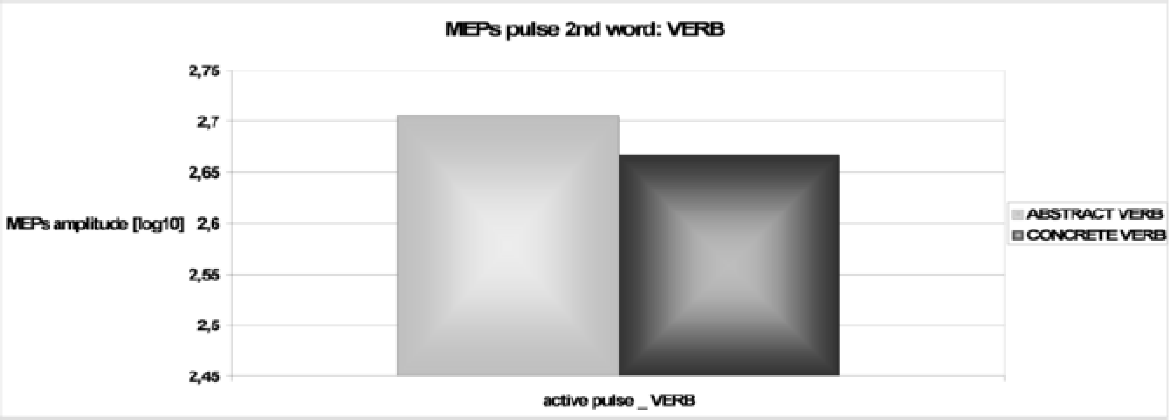


Figure 1. Normalized MEPs recorded after the TMS stimulation on the 2nd word. Peak-to-peak MEPs amplitude was greater for sentences containing Abstract Verbs than for sentences containing Concrete Verbs.

2.2 Analyses on MEPs after a pulse on 2nd word for both sensible and non-sensible sentences

As we found no effect of the kind of noun (abstract vs. concrete), in order to disambiguate the role of the kind of pulse (precocious, 250 ms, vs. delayed, 650 ms) and the kind of subsequent noun (determining a sensible vs. non-sensible combination) on the verb, in a further analysis we considered also the MEPs recorded from the FDI during non-sensible sentence processing (see Kocha et al., 2010). Normalized MEPs (after a pulse on 2nd word) were submitted to a 2 (Sentence: Sensible vs. Non-sensible) X 2 (Verb: Concrete vs. Abstract) X 2 (Noun: Concrete vs. Abstract) ANOVA. We conducted the analysis with participants as a random factor. We found a significant interaction between the Sentence and the Verb, F (1, 13) = 27.47, MSe = 0.001, $p < .001$: sensible sentences containing abstract verbs obtained greater peak-to-peak MEPs amplitude, M = 2.71, than sensible sentences containing concrete verbs, M = 2.67, post hoc LSD: $p < .0005$, see Fig. 2. Crucially, in the case of meaningless context we found an opposite pattern, that is greater peak-to-peak MEPs amplitude with Non-sensible Sentences containing Concrete (M = 2.69) rather than Abstract Verbs (M = 2.67, post hoc LSD: p < .05). Finally we found a three way interaction between the Sentence, the Verb and the Noun, F (1, 13) = 11.24, MSe = 0.001, p < .005: sensible sentences containing abstract verbs followed by abstract nouns obtained greater MEPs, M = 2.73, than sensible sentences formed by an abstract verbs plus a concrete noun, M = 2.69, post hoc LSD: p < .05; we did not find an analogous modulation for non-sensible sentences (p = .44).

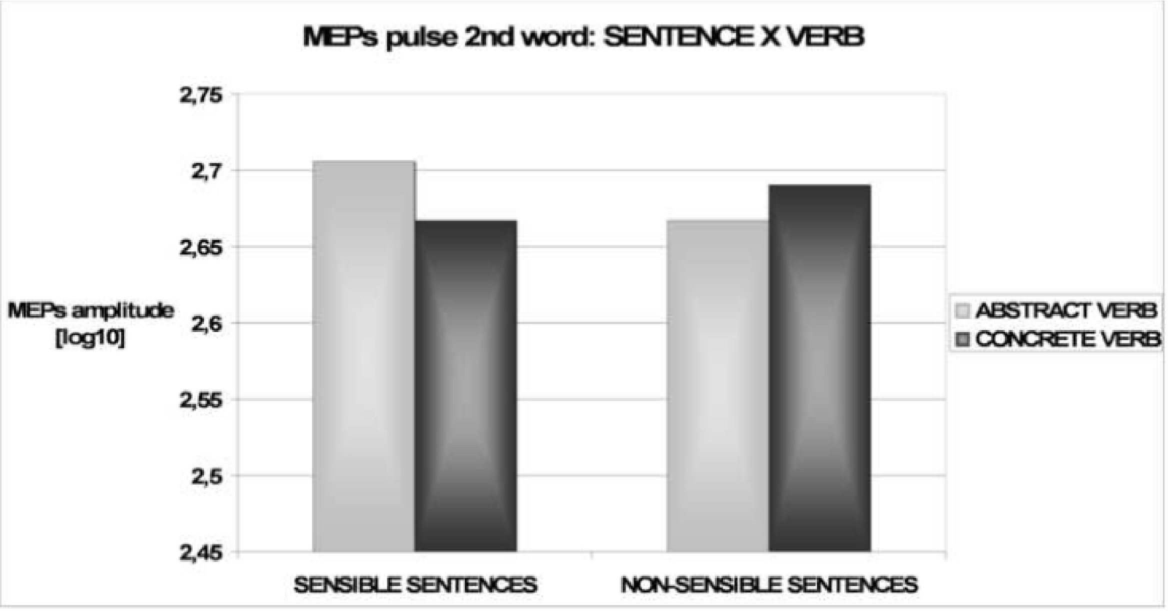


Figure 2. The interaction between the kind of Sentence and the Verb: Sensible sentences containing Abstract Verbs obtained greater MEPs amplitudes than Sensible sentences containing Concrete Verbs. We found an opposite pattern for Non-sensible sentences.

2.3 Analyses on accuracy and RTs

One participant was excluded from behavioral analyses due to the high percentage of unrecorded data (35.6%) because of difficulties with the response device. Percentages of errors were submitted to a 2 (Pulse: Active vs. Sham) X 2 (Verb: Concrete vs. Abstract) X 2 (Noun: Concrete vs. Abstract) X 2 (Stimulated Word: 1st word vs. 2nd word) ANOVA; we manipulated all variables within participants. Results showed a main effect of the Verb: participants made more errors with sentences containing Abstract Verbs (M = 0.59%) compared to sentences containing Concrete Verbs (M = 0.37%), F (1, 12) = 18.97, MSe = 0.141, p < .001. Analyses also showed a significant interaction between the Verb and the Noun, F (1, 12) = 19.71, MSe = 0.286, p < .001, basically due to the high number of errors in Abstract Verbs followed by Concrete Nouns (M = 0.81%) condition, that significantly differed from Abstract Verbs followed by Abstract Nouns (M = 0.38%, post hoc LSD: p < .001), Concrete Verbs followed by Concrete Nouns (M = 0.26%, post hoc LSD: p < .001) and Concrete Verbs followed by Abstract Nouns (M = 0.48%, post hoc LSD: p < .01) conditions. Finally we found an interaction between the Pulse, the Stimulated Word and the Verb, F (1, 12) = 7.44, MSe = 0.060, p < .05: in the Active Pulse condition with Sentences containing Abstract Verbs participants made more errors (M = 0.78%) when the pulse was delivered on the second word than on the first one (M = 0.56%, post hoc LSD: p < .01; the effect was not replicated for the control-sham condition, p = .12); we found no effect of the Stimulated Word Sentences containing Concrete Verbs (post hoc LSD: p = .28; control-sham condition, p = .58).

Before performing analyses on response times all incorrect responses were eliminated (3.57 %). Response times (ms) were submitted to a 2 (Pulse: Active vs. Sham) X 2 (Verb: Concrete vs. Abstract) X 2 (Noun: Concrete vs. Abstract) X 2 (kind of Stimulated Word: 1st word vs. 2nd word) ANOVA, with all variables within participants. Results showed a main effect of the kind of Noun, F (1, 12) = 5.05, MSe = 3966.197, p < .05: sentences containing

Abstract Nouns (M = 556.32 ms) were processed faster than sentences containing Concrete Nouns (M = 575.94 ms). This result appears to be due to the very slow response times obtained with Abstract Verbs plus Concrete Nouns combinations. Indeed, due to our particular paradigm, we collapsed verb and noun RTs focusing on sentences. As a result Concrete Noun processing turned out to be slower than Abstract Noun processing because the timing reflected not only the process of noun comprehension, but also the process of previous verb comprehension, as well as a possible delay caused by the switching cost. We will not discuss this result as it is partially explained by the interaction between Verbs and Nouns: analyses showed a significant interaction between the Verb and the Noun, F (1, 12) = 36.86, MSe = 1740.424, $p < .0001$, as participants were faster with congruent combinations (AA: M = 546.29 ms; CC: M = 550.84 ms) than with the mixed ones (AC: M = 601.04 ms; CA: M = 566.34 ms, post hoc LSD: $p < .05$). The advantage of congruent over mixed combinations replicates results found by Scorolli et al (2011) in a behavioral task employing the same paradigm. Additionally, post hoc LSD showed that participants employed the slowest response times with Abstract Verbs combined with Concrete Nouns: the modality switching (from concrete to abstract, or vice-versa) determines a delay; this delay is larger in case of sentences containing Abstract rather than Concrete Verbs. Most crucially we also found a three way interaction between the Pulse, the Stimulated Word and the Verb, F (1, 12) = 4.77, MSe = 3012.27, $p < .05$: when the pulse was delivered on the 1st word (verb), sentences containing Concrete Verbs (M = 538.55 ms) were processed faster than sentences containing Abstract Verbs (M = 576.11 ms, post hoc LSD: $p < .05$); in the control-sham condition we found no effect of the Verb ($p = .64$). The time latencies for sentences containing concrete verbs (M = 561.46 ms) and sentences containing abstract verbs (M = 565.65 ms) did not differ when the pulse was delivered on the 2nd word (noun, $p = .10$); not effect of the kind verb was found for the sham condition ($p = .16$). This result clearly argues in favor of a greater activation of the motor system during Concrete Verbs processing in case of TMS pulse.

To better understand our results we performed two further separated analyses focusing on the Sham Condition and on the Active pulse condition: for both the analyses, response times (ms) were submitted to a 2 (Verb: Concrete vs. Abstract) X 2 (Noun: Concrete vs. Abstract) X 2 (Stimulated Word: 1st word vs. 2nd word) ANOVA, with all variables manipulated within participants. In the Sham Condition analysis we found only a significant interaction between the Verb and the Noun, $F(1, 12) = 15.28$, $MSe = 2476.217$, $p < .005$: participants were faster with congruent combinations (AA: $M = 551.06$ ms; CC: $M = 554.52$ ms) than with the mixed ones (AC: $M = 601.85$ ms; CA: $M = 579.93$ ms). The interaction between the Verb and the Noun was significant also in the separated analysis on Active Stimulation Condition, $F(1, 12) = 6.50$, $MSe = 4121.074$, $p < .05$: participants were faster with congruent combinations (AA: $M = 541.52$ ms; CC: $M = 547.26$ ms) than with the mixed ones (AC: $M = 600.24$ ms; CA: $M = 552.75$ ms). Most crucially in this analysis we also found a significant interaction between the Verb and the kind of Stimulated Word, $F(1, 12) = 4.90$, $MSe = 1477.771$, $p < .05$: sentences containing Concrete Verbs were processed faster when the pulse was delivered on the 1st word (verb, $M = 538.55$ ms) than on the 2nd (noun, $M = 561.46$ ms, post hoc LSD: $p < .05$, see Fig. 3); conversely with sentences containing Abstract Verbs we found no effect of the Stimulated Word ($p = .35$).

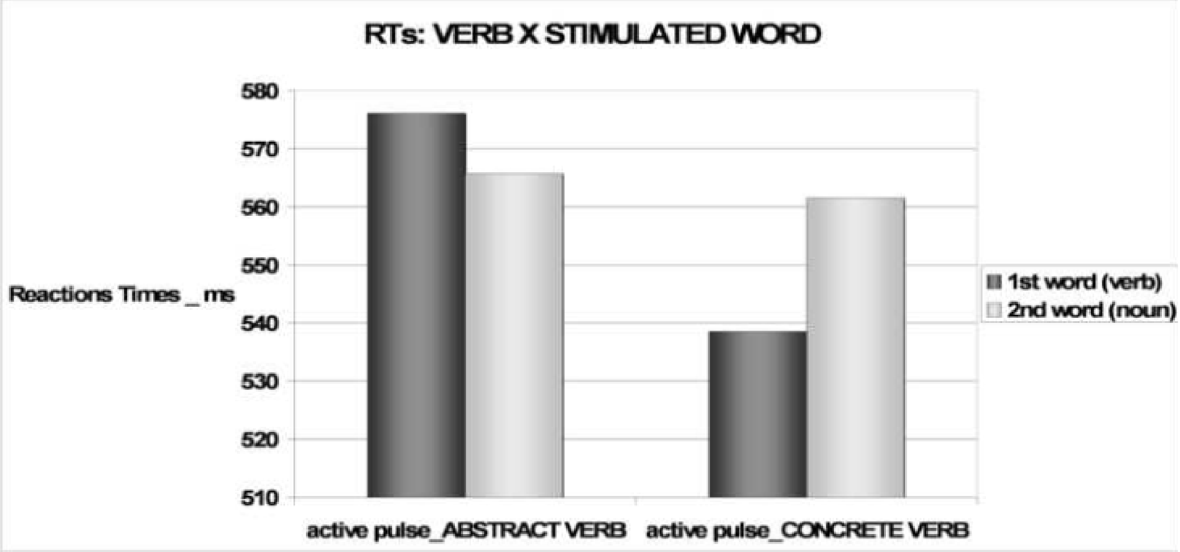


Figure 3. The interaction between the Verb and the kind of Stimulated Word: sentences containing Concrete Verbs were processed faster when the pulse was delivered on the 1st word than on the 2nd; with sentences containing abstract verbs no effect was found.

3. DISCUSSION

An important challenge of embodied theories is to explain the comprehension of abstract sentences. We performed a transcranial magnetic stimulation (TMS) study to explore the role of the left primary motor cortex during the processing of action-related and non-action-related verbs with nouns of graspable and non-graspable objects. Participants performed a sentence sensibility task. Single TMS pulses were delivered 250 ms after verbs vs. nouns presentation.

The first important result, supporting embodied theories, is that both concrete and abstract words modulate the activity of the motor system, as indicated by analyses on MEPs and suggested by results on RTs, even though this modulation involves different temporal windows. In addition, as predicted, our results suggest that concrete words activate the hand-related motor system in a more direct and straightforward way, while abstract words activate it in a different fashion. Both results are in line with a subset of embodied and grounded theories on abstract concepts and words processing. Indeed, they do not support embodied theories according to which concrete and abstract words do not differ in processing and representation. Rather, they support multiple representation views, and particularly the WAT proposal. We will now discuss the results that led us to this conclusion.

1. Precocious simulation with concrete verbs. In a previous study Pulvermüller et al. (2005) found a specific and precocious (150 ms) facilitatory effect of TMS subthreshold stimulation of the motor cortex on the action words processing. In their study, participants were presented with single words referring to leg (e.g., to kick) or hand-arm actions (e.g., to pick) and were asked to perform a lexical decision task. Leg words recognition was faster when TMS targeted the leg area than when TMS was delivered over upper limb representation; symmetrical results were obtained for hand-arm verbs. The results showed that the activation of motor and premotor areas modulates the processing of specific kinds of words, semantically related to the arm or the leg. Our study extends their results showing the temporal evolution of the language and action systems linkage in case of whole sentence processing (for a study on single verb processing using different temporal windows see Papeo et al., 2009; see also Liuzza et al 2011 for a study on positive and negative abstract and concrete sentences), using a language comprehension task (for a recent study on the functional anatomy of the language comprehension network see Turken et al., 2011). Analyses of MEPs after the pulse only on the first word do not allow us to draw any conclusion on Concrete Verbs processing. However the separate analysis we performed to contrast the kind of verb and the timings of TMS delivery showed that abstract verbs elicited greater peak-to-peak MEPs amplitude with a delayed pulse (650 ms) than with a precocious one (250 ms); crucially

concrete verbs presented an opposite pattern (sse also Candidi, Leone-Fernandez, Barber, Carreira and Aglioti, 2010).

Moreover our data on non-sensible sentences (pulse on the second word) are very informative. Indeed, we found that in case of meaningless context, that is when the whole sentence processing is broken as it is impossible to integrate the noun with the previous verb, motor activation is stronger for verbs referring to physical actions performed with the hand (greater peak-to-peak amplitudes with concrete than abstract verbs). It seems that when participants have to evaluate the sensibility of a sentence, as in the present task, they do not process the single words sequentially; rather the meaning emerges from the combination of words in sentences (see Pulvermüller, 2011), consistently with recent findings on single words formed by different morphological components (Rueschemeyer, Brass and Friederici, 2007). If the integration cannot be accomplished due to semantic constraints the comprehension process stops at the verb level. Together with the results on MEPs for sensible sentences, these results on both meaningless vs. meaningful sentences help us rule out a possible alternative explanation, that motor activity may increase merely as a function of task difficulty (e.g., Davis and Johnsrude, 2003; Fridriksson, et al., 2008). Analysis on the sentence sensibility judgment task (response latencies) showed consistent findings, as we found an advantage for sentences containing concrete rather than abstract verbs only when the pulse was delivered on the 1st word. Interestingly this effect seems to be very precocious, as the first pulse was delivered just 250 ms after the word presentation.

2. Verb-noun integration: late simulation with abstract verbs. As we presented verb-noun combinations, instead of single words, it is crucial to understand how the integration process of verbs and nouns takes place.

2a. Our results show that MEPs peak to peak amplitudes after a ‘delayed’ pulse (pulse on the second word, the noun) were greater with sentences containing abstract verbs than sentences containing concrete verbs. This result favors the hypothesis that abstract words (verbs) also activate the motor system (specifically, in our study, the motor system related to manual action; see Jirak et al, 2010), but this activation is delayed with abstract words than with concrete words, as suggested by separate analysis contrasting the kind of verb and the timings of TMS delivery, regardless of the noun (as it did not modulate the MEPs). The role of the precocious or delayed pulse and the context is disambiguated by further analysis we performed on both sensible and non-sensible sentences: crucially, this greater delayed involvement of the motor system in the case of sentences containing abstract verbs disappears with non sensible sentences (for which presumably the noun is not integrated with the verb). This suggests that simulation related to the semantic meaning of the sentence only occurs when the content makes sense and that this process leads to activation of the motor system.

We propose two possible explanations for this effect, relying on two different embodied views: (a) the motor simulation is also evoked by sentences containing abstract verbs, but this simulation occurs later than with sentences containing concrete verbs. This interpretation is consistent with a recent embodied theory that, similarly to WAT, proposes that multiple kinds of representation underlie knowledge, the Language and Situated Simulation Theory, LASS (Barsalou et al., 2008). According to LASS linguistic forms and situated simulations interact continuously, but while the linguistic system is mainly involved during precocious superficial linguistic processing, a deeper conceptual processing would be necessary for the operation of the simulation system (e.g., sensorimotor system activation; for consistent results, see also Louwerse & Connell, 2011). This proposal can account for the delayed activation of sentences containing abstract as opposed to concrete verbs, but LASS would predict a modulation of MEPs by the kind of noun. We found greater activation of the motor system for abstract verbs than concrete ones in MEPs after a pulse on the noun (400+250 ms). From 250 ms after the noun onset participants should have already processed the noun (Pulvermüller et al., 2005); so LASS would predict greater MEPs for graspable than for non-graspable objects combined with abstract verbs. Instead, we did not find any modulation of the noun; moreover, in the analyses of both sensible and non-sensible sentence we found an opposite pattern.

The second possible explanation (b) of this result supports the WAT proposal (Borghi and Cimatti, 2009; 2012). Due to their acquisition modality, concrete words evoke more manual information, while abstract words elicit more verbal information (Borghi et al., 2011). We can account for these results through arguing that concrete verbs activate precociously motor areas related to the hand, while abstract verbs activate precociously motor areas related to the mouth, as data on acquisition modality suggest (Borghi et al., 2011). The early activation of motor areas related to the mouth would have a delayed effect on motor areas related to the hand, due to their topological contiguity. The reason why MEPs modulation should be similar for both a direct effect (hand) and an indirect effect (mouth) might not seem straightforward. However, one could speculate that, in the temporal window of 250 ms, we might detect the hand related curve in its decreasing phase, while in the temporal window of 650 ms we might detect the curve describing the effect of the mouth on the hand areas during its increasing phase. On this basis, the signal that we detect at 650 ms could be a compound of mouth induced activation (abstract verbs) plus the activation determined by noun processing, that - resting in our measures - we cannot estimate, but that is reasonably different from zero (null activation). This interpretation is consistent with a study on visual, motor and abstract words by Kellenbach et al (2002): measuring event related potentials (ERPs). They found greater anterior positivity (lateral sites) activation with abstract words than with motor words starting from 300 ms; the effect lasted until 500 ms (centro-anterior sites). Later the effect became left lateralized (550-750 ms). Further results in line with our perspective have been found by Desai et al. (2010) with fMRI. Participants were presented with sentences of the form ‘‘I/You/We/They <verb > the <noun >’’ (e.g., “I throw the ball”) and had to evaluate their sensibility by pressing a key; they had to respond only to non sensible sentences. The sentences included either a motor (e.g., “grasp”), visual (e.g., “read”) or an abstract verb (e.g., “explain”, “allow”) combined with concrete and abstract nouns (e.g., “ball” vs. “method”). The results showed that abstract sentences, differently from motor and visual ones, strongly activated the superior/anterior temporal and inferior frontal areas. In line with WAT, this study on sentence processing suggests that the meaning of abstract words may be represented primarily through verbal associations with other words. The difference between Paivio’s view and embodied multiple representation views such as WAT is that, according to the last, both sensorimotor and linguistic information are crucial for both concrete and abstract words, even if the distribution of the two information sources is different. In our study the analysis on MEPs when the pulse was delivered on the second word indicates that also abstract verbs activated the manual motor system, even if it is unclear from the present study whether this activation of the manual system is the cascade effect of the involvement of the mouth areas (see below).

2b. Beyond the analysis on MEPs, the second main result on verb-noun integration is from reaction times analyses. We found an interaction between the Verb and the kind of pulse: sentences containing concrete verbs were processed faster than sentences containing abstract verbs when the TMS pulse was delivered on the presentation of the verb. We did not find any difference when the pulse was delivered on the noun. Consistently with our interpretation of MEPs, reaction times were faster when the hand related motor areas were directly involved (concrete verbs). The supposed indirect activation of hand areas by abstract verbs affected the MEPs but it did not last long enough, and probably was not strong enough, to affect response times. Finally, the interaction between the Verb and the kind of Noun is consistent with a recent cross-linguistic study (Scorolli et al., 2011) in which we found the same advantage a. for compatible combinations, and, b. within the mixed combination, when the concrete word preceded the abstract word, regardless of its grammatical class (see Paivio, 1965).

Overall our results seem to indicate that while sentences containing concrete verbs imply a direct precocious activation of the hand related motor system, the activation of the same system is delayed in the case of sentences containing abstract verbs. The processing of abstract verbs could precociously engage mouth related motor areas, that later affect the contiguous areas (hand areas).

However, the present evidence does not allow for disambiguation between two alternative explanations: (1) abstract words have a weaker grounding in the sensorimotor system; (2) abstract words are processed in an alternative route, maybe in the premotor cortex, with involvement from mouth related motor areas. Integrating these results with those recently obtained in a study on novel words acquisition (Borghi et al., 2011) we lean towards the second hypothesis.

As hypothesized by the WAT proposal, mouth areas could be crucial for abstract word processing. In thinking about the acquisition of a concrete word, such as “pencil”: the acquisition simply requires a person to use the label while indicating the right referent. The acquisition of a concept-word like “democracy”, instead, implies the presence of somebody explaining the word meaning, using language. This experience is still a bodily experience but the contribution of the social dimension is more relevant to acquisition. In addition, in this experience language is not only the counter part of an external referent but is a tool that allows us to acquire more complex meanings, a powerful means of collecting a variety of bodily and situational experiences.

4. EXPERIMENTAL METHOD

4.1 Participants

16 students (7 men and 9 women; mean age = 27.44 years; s.d. = 1.93) attending the University of Bologna took part in the study. All were native Italian speakers, right-handed and all had normal or corrected-to-normal vision. Before starting the experimental session, the experimenters assessed their general health status with a brief interview: none of them were reported evidence for neither neurological or psychiatric diseases, nor contraindications related to single-pulse TMS procedure. Then participants were provided with a detailed explanation about the procedure, contraindications and risks of the experiment (Wessermann, 1998). To begin the experiment participants had to confirm their voluntary participation by written consent. The study was approved by the local ethics committee (Department of Psychology, University of Bologna). All participants received compensation for their participation.

4.2 Transcranial Magnetic Stimulation and EMG recording

As an index of corticospinal excitability, we recorded motor-evoked potential (MEPs). MEPs induced by TMS were recorded from the right first dorsal interosseus muscle (FDI, in the region of the index finger) by means of a Biopac Student Lab MP36 electromyograph (Biopac Systems, Inc, U.S.A.). EMG signals were band-pass filtered (20 Hz–2.5 kHz, sampling rate fixed at 10 kHz), digitized and stored on a computer for off-line analysis. Pairs of silver/silver chloride surface electrodes were placed over the muscle belly (active electrode) and over the associated joint or tendon of the muscle (reference electrode). A circular ground electrode with a diameter of 30 mm was placed on the internal bone of the right elbow. Single-pulse TMS was applied to the left M1, using a Magstim Rapid 2 stimulator (Magstim, Whitland, Dyfed, U.K.) connected to a figure-of-eight coil (70 mm in diameter). The coil was moved over the left hemisphere to determine the optimal position from which maximal amplitude MEPs were elicited in the FDI muscle. The optimal scalp position for the induction of MEPs with the maximum amplitude in the right FDI muscle was individuated for each participant. The coil rested tangential to the scalp with the handle pointing backwards and laterally at a 45° angle away from the midline. The target site was marked with a drawing pen on a cap applied on participants' head, and the coil was maintained in position by the experimenters. The intensity of magnetic pulses was set at 120% of the resting motor threshold (rMT), which is the minimum intensity of output required to produce MEPs with amplitude of at least 50 μV in the FDI muscle with 50% of probability (Rossini et al., 1994). The absence of voluntary contraction was continuously verified visually and, prior to the recording session, through auditory monitoring of the EMG signal.

4.3 Linguistic materials

Stimulus materials consisted of word pairs composed of a transitive verb and a concept noun. We used 28 quadruplets, thus 112 sensible sentences. Each quadruplet was constructed by pairing a concrete verb (e.g. to grasp) with a concrete noun (e.g. a flower) or an abstract noun (e.g. a concept); and by pairing an abstract verb (e.g. to describe) with the previously used concrete and abstract noun. We defined Concrete Nouns as nouns referring to graspable objects and Concrete Verbs as verbs referring to physical actions (Taylor, 1977; Vendler, 1957) performed with the hand. We defined Abstract Nouns as nouns that do not refer to graspable objects and Abstract Verbs as verbs expressing mental processes, with no reference to a physical object (Taylor, 1977; Vendler, 1957). To select the 28 critical quadruples from 48 ones, we asked twenty Italian students to judge the familiarity of each sentence and with what degree of probability they would use each sentence. We then selected the quadruples with highest scores in both ratings and with lowest scores in the standard deviations (for a detailed description of the materials' selection see Scorolli et al., 2011).

In order to further test if the selected pairs differed in written frequency of use we utilized the research engine "Google": we checked the number of occurrences of each verb-noun pair, by using quotations marks (Page et al., 1998; Griffiths et al., 2007; Sha, 2010). The obtained frequencies were submitted to a 2 (Noun: Concrete vs. Abstract) X 2 (Verb: Concrete vs. Abstract) ANOVA. Crucially, we did not find any significant effect (all $p_s \geq .41$). The establishment of control on written frequency allowed us to exclude that processing differences rest on different degrees of association between the words pairs used in the quadruples. Finally we selected 112 non-sensible sentences, that is sentences in which the actions described by the abstract (e.g. to suspect) or concrete (e.g. to eat) verbs were not suitable for the abstract (e.g. the freedom) or concrete (e.g. a pen) nouns that followed the verb (non-sensible sentences). Due to the particular kind of paradigm it was impossible to balance sentences for word length and number of syllables. However, this should not represent a problem, given that our main hypotheses pertain to the interactions.

4.4 Procedure

The experiment was programmed using the EPrime (Psychology Software Tools, Inc, U.S.A) software to control sequence and duration of the presentation of the linguistic material, and to trigger TMS and EMG recording. Participants were asked to perform a sentence sensibility task: they were required to judge if sentences made sense or not. Participants focused on a fixation cross; after 1000 ms a verb appeared on the screen; after 500 ms the verb was substituted by a noun. The cut-off was set at 2500 ms from the noun onset. On conclusion of this cycle there was a pause; then the next trial began. Each trial lasted 8000 ms from start to finish, i.e. long enough to prevent interaction between consecutive TMS-pulses (Robertson et al., 2003). Participants were instructed to use the left foot – homolateral side with respect to TMS stimulation site – to respond. They were randomly assigned to one of two groups. Participants in the first group were asked to respond “yes” (= the combination makes sense) pressing the right pedal and “no” (= the combination doesn’t make sense) pressing the left pedal; participants in the other group were assigned the opposite mapping. Participants were instructed to keep their right arm/hand and head motionless and muscle relaxation was monitored throughout the entire experiment to check for involuntary movements.

Response times and errors were recorded using EPrime; the timer started from the noun presentation. The experiment consisted of four blocks of 112 items each (56 sensible sentences and 56 non-sensible sentences). For each block participants were alternatively delivered a TMS (2 blocks) or a sham (2 blocks) stimulation, randomly delivered 250 ms after the onset of the first word (verb) or of the second word (noun). The choice of the temporal window was motivated by electrophysiological evidence showing that starting from 250 ms motor words elicited greater negativity than both visual and abstract words (Kellenbach, Wijers, Hovius, Mulder, & Mulder, 2002). The order of the two stimulation conditions was counterbalanced across subjects. To mimic the TMS conditions (Robertson et al., 2003), in the sham stimulation conditions the same intensity of magnetic pulse was used, but a cylinder made of insulating material was located between the coil and the scalp surface. Each sentence was presented twice, so we collected 14 observations for each experimental condition. 224 motor evoked potentials (MEPs) were obtained from each participant, one magnetic stimulus being applied for each item (the pulses delivered during the two sham-blocks did not elicit MEPs). The four kinds of sentences were presented in random order within each block, with a short pause after 28 items.

At the end of the experiment participants were debriefed. Since none of them was previously exposed to TMS, they reported that they had attributed the differences in the peripheral effects intensity in the sham and TMS conditions to different pressures applied on the scalp by the two experimenters.

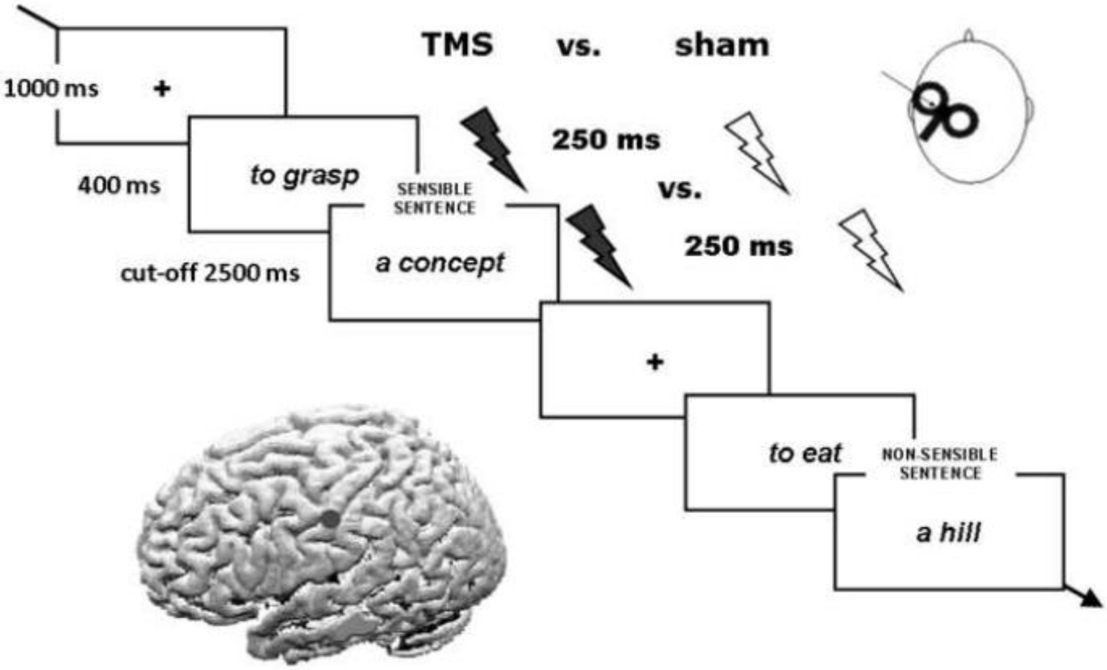


Figure 4. The figure shows the experimental paradigm. The coil was moved over the left hemisphere to determine the FDI representation in the primary motor cortex.

Acknowledgments

Thanks to Michele Marzocchi for the useful script to compute screen for MEPs >= 50 μV. Thanks to Alessio Avenanti for discussions and useful suggestions on this paper and to Roberto Bolzani for statistical advice. Thanks to Kate Burke for revision of the English text. This work was supported by the European Community, project ROSSI: Emergence of communication in RObots through Sensorimotor and Social Interaction (Grant agreement n. 216125).

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577–660.

Barsalou, L.W. (2003). Situated simulation in the human conceptual system. Language and Cognitive Processes, 18, 513-562.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.). Symbols, embodiment, and meaning (pp. 245–284). Oxford, U.K.: Oxford University Press.

Binder, J.R., Westbury, C.F., McKiernam, K.A., Possing, E.T., & Medler, D.A. (2005). Distinct Brain Systems for Processing Concrete and Abstract Concepts. Journal of Cognitive Neuroscience, 17, 905-917.

Bird, H., Franklin, S, & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. Behavior Research Methods, Instruments & Computers. Special Issue, 33, 73-79.

Borghi, A.M. (in press). Action language comprehension, affordances and goals. In Y. Coello & A. Bartolo (Eds). Language and action in cognitive neuroscience. Psychology Press.

Borghi, A.M., & Cimatti, F.(2009). Words as tools and the problem of abstract words meanings. In N. Taatgen & H. van Rijn (eds.). Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp. 2304-2309). Amsterdam: Cognitive Science Society.

Borghi, A.M., & Cimatti, F.(2010). Embodied cognition and beyond: Acting and sensing the body. Neuropsychologia, 48, 763-773.

Borghi, A.M., & Cimatti, F. (2012). Words are not just words: the social acquisition of abstract words. Italian Journal of Philosophy of Language, 5, 22-37. doi: 10.4396/20120303.

Borghi, A.M., Flumini, A., Cimatti, F., Marocco, D. & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. Frontiers in Psychology**,** 2:15. doi: 10.3389/fpsyg.2011.00015.

Borghi, A.M. & Pecher D (2011). Introduction to the special topic Embodied and Grounded Cognition. Frontiers in Cognition, 2:187. doi: 10.3389/fpsyg.2011.00187.

Candidi, M., Leone-Fernandez, B., Barber, H.A., Carreiras, M., & Aglioti, S.M. (2010). Hands on the future: facilitation of cortico-spinal hand-re presentation when reading the future tense of hand-related action verbs. European Journal of Neuroscience, 32, 677–683.

Casasanto, D. (2009). Embodiment of Abstract Concepts: Good and bad in right- and left-handers. Journal of Experimental Psychology: General 138, 351-367.

Clark (2007). Curing Cognitive Hiccups: A Defense of the Extended Mind. Journal of Philosophy 104,163-192.

Davis M. H., Johnsrude I. S. (2003). Hierarchical processing in spoken language comprehension. Journal of Neuroscinece, 23, 3423–3431.

Decety, J., & Grèzes, J. (2006). The power of simulation : Imagining one's own and other's behavior. Brain Research, 1079, 4-14.

Decety, J., & Ingvar, D.H., (1990). Brain structures participating in mental simulation of motor behavior: a neuropsychological interpretation. Acta Psychologica, 73, 13–24.

Desai, R.H., Binder, J.R., Conant, L.L., & Seidenberg, M.S. (2010). Activation of Sensory-Motor Areas in Sentence Comprehension. Cerebral Cortex, 20, 468-478.

Dove, G. (2009). Beyond Conceptual Symbols. A Call for Representational Pluralism. Cognition, 110, 412-431.

Dove G (2011) On the need for embodied and dis-embodied cognition. Frontiers in Psychology 1:242. doi: 10.3389/fpsyg.2010.00242.

Fridriksson J, Moss J, Davis B, Baylis G, Rorden C. (2008) Motor speech perception modulates the cortical language areas. Neuroimage, 41: 605–613.

Gallese, V. (2009). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. Psychological Research, 73, 486-98.

Gibbs, R., & Steen, G. (1999). Metaphor in cognitive linguistics. Amsterdam: John Benjamins.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. Journal of Memory & Language, 43, 379-401.

Goldman, A., De Vignemont, F. (2009). Is social cognition embodied ? Trends in Cognitive Science, 13, 154-9.

Griffiths, T.L., Steyvers, M., & Firl, A. (2007). Google and the Mind. Predicting Fluency With PageRank. Psychological Science, 18, 1069-1076.

Jirak, D., Menz, M., Buccino, G., Borghi, A.M., & Binkofski, F. (2010). Grasping language. A short story on embodiment. Consciousness and Cognition, 19, 711-720.

Kellenbach, M.L., Wijers, A.A., Hovius, M., Mulder, J., & Mulder, G. (2002) Neural differentiation of lexico-syntactic categories or semantic features? Event-related potential evidence for both. Journal of Cognitive Neuroscience,14, 561-577.

Kocha, G, Versacea, V., Bonnì, V.S., Lupoa, F., Lo Gerfoa, E., Oliveri, M., & Caltagirone, C. (2010). Resonance of cortico–cortical connections of the motor system with the observation of goal directed grasping movements. Neuropsychologia, 48, 3513–3520.

Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press.

Liuzza M.T., Candidi M., Aglioti S.M. (2011). Do not resonate with actions: sentence polarity modulates cortico-spinal excitability during action-related sentence reading. PLoS One, 6(2):e16855.

Louwerse, M.M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. Cognitive Science, 35, 381-398.

Louwerse, M.M., & Jeuniaux, P. (2009). Language comprehension is both embodied and symbolic. In M. de Vega, A. Glenberg, & A.C. Graesser (Eds.), Symbols, embodiment, and meaning: A debate (pp. 309-326). Oxford: Oxford University Press.

Mirolli, M., & Parisi, D. (2011). Towards a Vygotskyan cognitive robotics: The role of Language as a Cognitive Tool. New Ideas in Psychology, 29: 298-311. doi:10.1016/j.newideapsych.2009.07.001.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the web (Tech. Rep.). Stanford, CA: Stanford Digital Library Technologies Project.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 4, 32-38.

Paivio, A. (1986). Mental representations: A dual coding approach. New York: Oxford University.

Paivio, A. (1991). Dual coding theory: retrospect and current status. Canadian Journal of Psychology 45, 255-87.

Paivio, A., Yuille, J. C. & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 76, 1-25.

Papeo, L., Vallesi, A., Isaja, A., Rumiati, R.I., (2009). Effects of TMS on Different Stages of motor and Nonmotor Verb Processing in the Primary Motor Cortex. PLoS ONE, 4, e4508.

Pecher, D., Boot, I., & van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. Ross (Ed.). The Psychology of Learning and Motivation, vol. 54 (pp. 217-248). Burlington: Academic Press.

Penfield W, & Rasmussen T (1950). The cerebral cortex of man. New York: Macmillan.

Pulvermüller, F., Hauk, O., Nikulin, V., & Ilmoniemi, R.J. (2005**).** Functional links between motor and language systems. European Journal of Neuroscience, 21, 793-797.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. J. (2005). Brain signatures of meaning access in action word recognition. Journal of Cognitive Neuroscience, 17, 884-892.

Pulvermüller, F. (2011). Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge. Journal of Neurolinguistics.

Robertson, E.M., Théoret H., Pascual-Leone A. (2003). Studies in cognition: the problems solved and created by transcranial magnetic stimulation. Journal of Cognitive Neuroscience, 15, 948–960.

Rodríguez-Ferreiro, J., Gennari, S.P., Davies, R., & Cuetos, F. (2010). Neural Correlates of Abstract Verb Processing. Journal of Cognitive Neuroscience 23, 106–118.

Rossini, P.M., Barker, A.T., Berardelli, A., Caramia, M.D., Caruso, G., Cracco, R.Q., Dimitrijevic, M.R., Hallett, M., Katayama, Y., Lucking, C.H., et al. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. Electroencephalogr Clin Neurophysiol, 91, 79-92.

Rueschemeyer, S.A., Brass, M., & Friederici, A.D. (2007) Comprehending Prehending: Neural Correlates of Processing Verbs with Motor Stems. Journal of Cognitive Neuroscience 19:5, pp. 855–865.

Schwanenflugel, P. (1991) Contextual constraint and lexical processing. In G. B. Simpson (Ed.), Understanding word and sentence. Amsterdam: Elsevier.

Scorolli, C., Binkofski, F., Buccino, G., Nicoletti, R., Riggio, L., & Borghi, A. (2011). Abstract and concrete sentences, Embodiment and languages. Frontiers in Psychology 2:227. doi: 10.3389/fpsyg.2011.00227

Sha, G. (2010). Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. Computer Assisted Language Learning, 23, 377-393.

Shallice, T, & Warrington, E.K. (1975). Word recognition in a phonemic dyslexic patient. Quarterly Journal of Experimental Psychology, 27, 187–199.

Simmons, W.K., Hamann, S.B., Harenski, C.N., Hu, X.P., & Barsalou, L.W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. Journal of Physiology – Paris, 102, 106-119.

Taylor, B. (1977). Tense and continuity. Linguist Philosophy, 1, 199–230.

Turken, A. U., & Dronkers, N.F. (2011). The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. Frontiers in Systems Neuroscience, 5

Tylèn, K., Weed, E., Wallentin, M., Roepstorff, A. & Frith, C.D. (2010). Language as a Tool for Interacting Minds. Mind & Language, 25, 3-29.

Vendler, Z. (1957). Verbs and times. Philosophical Review, 66, 143–160.

Wang, J., Conder, J.A., Blitzer, D. N. & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A Meta-Analysis of Neuroimaging Studies. Human Brain Mapping, 31, 1459-68.

Warrington, E.K. (1975): The selective impairment of semantic memory. Quarterly Journal of Experimental Psychology, 27, 635–657.

Wessermann, E.M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the international workshop on the safety of repetitive transcranial magnetic stimulation. Electroencephalogr Clin Neurophysiol, 108, 1–16.

Zwaan, R.A. (2004). The immersed experiencer: toward an embodied theory of language comprehension. In: B.H. Ross (Ed.), The Psychology of Learning and Motivation, 44, 35-62. New York: Academic Press.

APPENDIX II

Scientific project submitted to the post-doctoral fellowship granted by the Fyssen Foundation

(deliberation: 15th July 2012)

INVESTIGATING THE NEURAL DYNAMICS OF HUMAN SOCIAL CONFORMITY SCIENTIFIC PROJECT

Prior expectations refer to the set of information on which individuals rely to make decisions in noisy or uncertain situations. These expectations are derived from prior knowledge that may originate from the past experience of the person (through expertise or learning of statistical regularities), from her intuitive theories, as well as from contextual information surrounding an action scene [1]. During my PhD, I participated in a series of studies that investigated how different types of prior information (i.e. biomechanical versus probabilistic priors) alter the perception and the prediction of goal-directed actions [2,3]. In a theoretical paper, I proposed that biomechanical priors (a class of priors based on a sensori-motor estimation of action costs)reduce the space of behavioral alternatives one may perform in a given situation and thereby directly influence the social learning of new skills and favor the emergence of stable behavioral traditions in a population [4]. Furthermore, using the technique of evolutionary individual-based modeling, we demonstrated that stable behavioral patterns can emerge at the population level via both high-fidelity and low-fidelity social learning mechanisms, as long as the latter are coupled with a highly constrained behavioral repertoire (e.g., constrained by biomechanical priors) [5]. The two-year project presented here naturally follows from these previous studies and extends on them by proposing to investigate another form of priors, namely expectations generated by social environments.

Two kinds of social conformity

Changing one’s behavior to adopt the behavior exhibited by a majority of peers is a form of social influence that has been termed *social conformity* [6,7]. The modification of an individual’s judgment under the pressure of a group was first demonstrated by Asch in a series of seminal experiments [8,9]. In these experiments participants were asked to estimate and make judgments about the relative lengths of line segments. Although these tasks were perceptually simple, participants frequently gave the wrong answers when a group of peers was also giving the wrong answer. Following Asch’s work, social psychologists agreed on a dual explanation of conformity split into *informational conformity* and *normative conformity* [10,11]. It has been proposed that the influence of these two types of conformity differs according to the number of sources in the influence group and the amount of uncertainty during decision-making [12,13].

Informational conformity helps an individual to pick up nonsocial information from behaviors displayed by other group members, and is guided by the need to maximize performance (e.g., being accurate in discriminating sensory stimuli). Informational conformity is particularly salient in private contexts – i.e., when the individual’s decisions are unseen by the group. In such contexts, the probability of an individual showing conformity is positively correlated with the amount of subjective uncertainty; the more uncertain a subject is the more likely she/he is to pick up information from the group in order to reduce the level of uncertainty. Furthermore, the effect of group size on the probability of conforming decreases with increasing group size: a minimum number of influence sources is assumed to be sufficient to reduce the individual’s uncertainty [12,13].

In contrast, *normative conformity* occurs when an individual picks up social information from the behavior displayed by other group members. Normative conformity biases are most often expressed in public contexts – i.e., when the individual's decisions are seen by the group. Such biases occur when individuals adapt their behavior to the group's expectations in order to maximize their own social outcomes (e.g., maintaining social cohesion or avoiding social exclusion). For this reason, the probability of conforming in a normative context depends less on subjective uncertainty or the reliability of the information provided by the group and more on social approval. Furthermore, the effect of the group size on the probability of conforming increases as a function of the group size: increasing the number of people giving feedback increases the weight of social information [12,13].

The neuroscience of social conformity

We intend to study whether informational and normative biases engage distinct cerebral processes during the treatment of social information which, in turn, differentially influence perceptual decisions.

The existing neuroscientific literature suggests that social information can be processed similar to any other type of rewarding stimulus [14,15]. For example, it has been shown that when the individual and group opinions conflict, the brain generates neural signals in the anterior cingulate cortex (a region in the posterior medial frontal cortex) and the ventral striatum that share common features with the prediction error signal carried by the dopaminergic system and observed during reinforcement learning [16,17]. Detection of conflicting social feedback appears to be processed by an individual as the detection of an erroneous action outcome and the the magnitude of subsequent behavioral adjustments is correlated with the magnitude of the 'error signal'. Furthermore, the strength of this 'prediction error' signal is correlated with the amplitude of the conflict, and is predictive of the individual's propensity to change future decisions to conform to the group [18]. Together, these results indicate that the encoding of social priors as a specific class of reward generates a top-down signal in frontal regions that influences complex decisions. Moreover, the likelihood of adjusting decisions to match those of the influence group is increased when the target stimuli are initially judged as uncertain. This suggests that the influence of social priors may be particularly strong when an individual has to make decisions in uncertain situations – i.e., when the available information does not sufficiently constrain the number of potential alternatives.

Berns et al. (2005) [19] used neuroimaging to demonstrate that, in a simple mental rotation task, when perceptual judgments conflicting with the group were followed by conformity, a functional change occurred in an occipital-parietal network known to be involved in mental rotation. This result suggests that – just as for other types of prior information [20,21] – social information processed in frontal regions of the brain generates prior expectations that bias low-level perceptual decisions operating in sensory cortices.

It is important to point out that all the above-mentioned studies differed in terms of the social context or the amount of subjective uncertainty. For example, Berns et al. (2005) were the only ones who simulated realistic individual-to-group interactions in which participants' responses were fully visible to other group members. Thus, unlike the other studies, their task context was appropriate for the emergence of a normative conformity bias (see above).

Whether or not this was crucial for the observed results is unclear, however, given that neither the context nor the amount of subjective uncertainty was systematically varied.

Without systematically varying context and perceptual uncertainty two fundamental questions in the social conformity realm will remain unanswered. First, how does the brain process social information according to i) the informational (private) and normative (public) context in which the decision is taken, ii) the number of sources in the influence group that provide feedback, and iii) the perceptual uncertainty during decision-making? Second, do these factors have a direct impact on brain mechanisms responsible for high-level and low-level decision-making? The aim of the research proposed here is to address these questions.

EXPERIMENTAL OUTLINE

EEG EXPERIMENT

Objectives. Using behavioral and neuroimaging methods, **the aim of the present project is to investigate whether, and how, expectations about information provided by a group of peers influence the neural mechanisms of perceptual decision-making.** We will use a protocol in which the participants' propensity to conform to group opinion (i.e., the weighting of social priors) will be manipulated within an ecological context of social influence. In this protocol, healthy adults are required to categorize facial expressions after having received computer-generated feedback simulating the responses of three bogus influence sources (see fig.1). The experiment will be composed of two separate tasks performed by a group of twenty healthy volunteers. In the *informational conformity task* participants will receive social feedback from three influence sources but will not directly confront these sources. This 'private' context should favor conformity to social feedback for informational motives. In the *normative conformity task* participants will receive the same social feedback from the three influence sources. In this task, however, they will direct confront the sources and will simultaneously perform the task with them. This 'public' context should favor conformity to social feedback for normative motives.

In each task we will vary i) the amount of perceptual uncertainty elicited by the target stimuli (ambiguous versus unambiguous faces) as well as ii) the distribution of social feedback (0, 1, 2 or 3 sources conflicting with the perceptual evidence elicited by the target stimulus). While participants perform the task we will use EGG to measure four well-known event-related potentials (ERPs). The feedback-related negativity (N_{FB}) reflects the prediction error signal that occurs during the monitoring of erroneous action outcome and predicts behavioral adjustment [22]. The feedback-related positivity (P_{FB}) is modulated by negative feedback that provides information relevant to the task goal [23]. The early posterior negativity (EPN) and the late positive potentials (LPP) are involved in the perceptual processing of facial expressions [24]. N_{FB} and P_{FB} reflect modulations of neural signal in the medial frontal region, and will be used to predict changes in decisional strategy. EPN and LPP assess modulations of neural signal in sensory and associative cortices, and will be used to predict changes in perceptual decision-making.

Hypotheses. We hypothesize that **social information provided by a group can affect perceptual decision making by changing neural activity in frontal areas and by exerting**

backward influences on sensory areas in which perceptual decisions are processed. We expect this effect to be modulated by perceptual uncertainty, the distribution of social feedback, as well as the informational and normative context of the task. More specifically, we expect social information to be processed as a rewarding stimulus when the need for social approval overtakes the need to maximize performance (normative context), in particular when feedback from the three sources reaches consensus and conflicts with perceptual evidence provided by unambiguous faces. This could result in a change in high-level decisional processes operating in medial frontal regions (evident as an increase in N_{FB} amplitude) without any alteration in low-level perceptual decisional processes (shown by no modulation of EPN or LPP). In contrast, we expect social information to be processed just like any other type of information when perceptual uncertainty is high (ambiguous faces) and when maximizing performance overtakes the need for social approval (informational context), and this, independently of whether social feedback reaches consensus. This could result in a change in high-level decisional processes (shown by an increase in P_{FB} amplitude) as well as a change in low-level decisional processes operating in sensory cortices (shown by a modulation of EPN or LPP).

Participants. Participants will be tested at the 'Laboratoire de Psychologie de la Perception' at the Paris Descartes University (Paris, France). The same twenty healthy volunteers will perform both informational and normative conformity. All participants will complete the self-monitoring scale of interpersonal influence [25], the Hare Psychopathy Checklist-Revised [26], and the State-Trait Anxiety Inventory [27].

Methods. Informational and normative tasks will be separated by a minimum of two weeks. Taking into account the constraints associated with EEG (long preparation time, the influence of fatigue on EEG signals, etc.), we will run each task in two identical testing sessions lasting approximately 45 min each and separated by 24 hours. In each 45 min session we will present 350 computer-generated male faces extracted from the free database developed by [28]. Subjects will be required to judge whether the face is friendly or threatening.

General procedure. A typical trial consists of two presentations of the same face stimulus: one before and one after the social feedback (see [18] for a similar procedure). A first screen will present a single face stimulus for 1000ms during that the participant is required to carefully observe (first presentation = observation phase). Then three pictures will appear at the center of the screen representing the influence sources. In the next screen the participant sees the decisions of the influence sources for the face presented on the first screen. After 1000ms the social feedback is followed by a black screen for between 200 and 800ms. The next screen then presents the face stimulus that was previously observed and judged by the influence sources (second presentation = decision phase) and participants will have 1000ms to make their own decision by pressing one of two computer keys corresponding to 'friendly' (green key) and 'threatening' (red key). After their decision they will be required to indicate their level of confidence on a 5 point scale by moving a cursor to the desired value. An intertrial interval will then be presented for a variable duration of 1500 to 3000ms.

Manipulating the informational and normative context. For the informational conformity task, participants will be informed that they will participate in the pilot rating of a new face stimuli database. The experimenter will clearly explain to participants that their responses will remain strictly private and anonymous (private context, see fig.1a). Participants will be told that, before answering, they will see a screen showing the decisions of three other participants who have already performed the task. After seeing the group’s decision, the participants will be required to give his/her own decision about the same stimulus.

The normative conformity task is identical to the informational conformity task, except that we will simulate realistic on-line interactions between the participant and the three influence sources (public context, see fig.1b). Prior to running the experiment, participants will meet the three sources (members of the lab). The participant and the three confederates will be informed that they will simultaneously perform the task, in different rooms, connected by a computer network. We will inform the participant that the three confederates will make their decisions after the first exposure to the stimulus, while she/he will make her/his decisions after the second exposure. A webcam attached to the computer screen will capture the participant’s and the three confederates behaviors, such that they can keep an eye on each others all along the task. Finally, we will inform the participant that each of her/his perceptual decisions will be communicated to the other members under the form of a colored pattern. Participants will be naïve to the real purpose of the tasks.

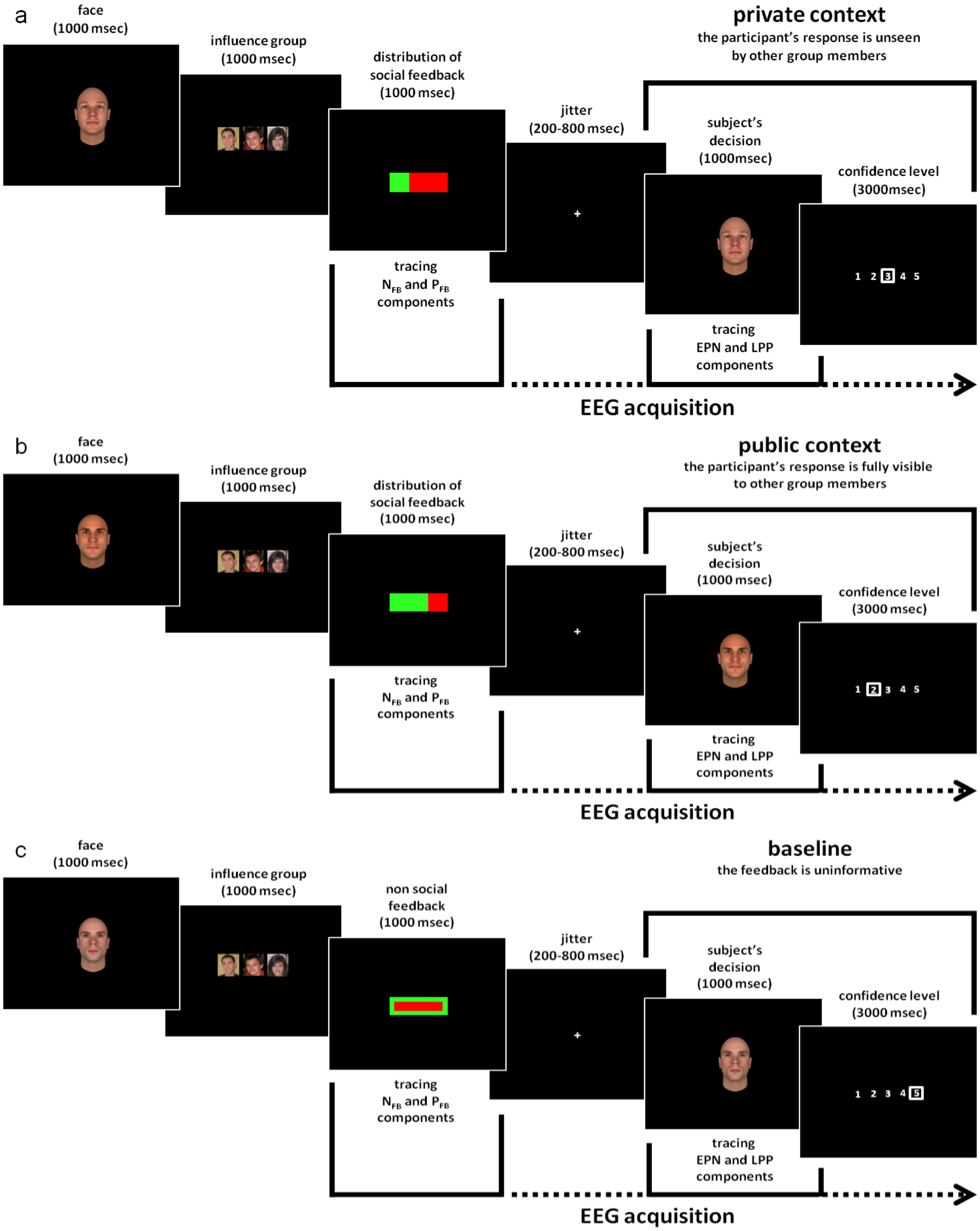


Figure 1a,b,c. Figure 1a represents a typical trial in the informational conformity task. In the example, the face is ambiguous (-1 value) but 2 out of 3 influence sources perceived it as threatening (rectangle predominantly coloured in red). Because the participant's initial percept is uncertain, she/he is expected to conform to the group majority (social information is used to reduce perceptual uncertainty and maximize performance) with a moderate to high level of confidence. Figure 1b represents a typical trial in the normative conformity task. Here, the face is unambiguously threatening (+2 value) but 2 out of 3 influence sources perceived it as friendly (rectangle predominantly coloured in green). Although the participant's initial percept is unambiguous, she/he is expected to conform to the group majority, but with a low level of confidence (social information overwhelms the perceptual evidence and is used to maximize social outcomes). In this task, the icons will display pre-recorded video clips of the three sources as if they were performing the run. Figure 1c represents the baseline condition for both tasks. Here, the feedback is not social as it does not give any information about the group's decisions. For both tasks, ERP components will be recorded during the third (N_{FB} and P_{FB}) and the fifth screen (EPN and LPP).

Manipulating perceptual uncertainty (see fig.2). The face stimuli will be distributed along a continuum, from friendly to threatening. During the tasks, we will manipulate the level of perceptual uncertainty by presenting unambiguous and ambiguous faces. Faces located at extremes of the continuum are 'unambiguous' (-2; +2) and elicit low perceptual uncertainty, faces located around the center are more 'ambiguous' (-1; +1) and elicit higher perceptual uncertainty. Of note is that faces labelled as 'ambiguous' do not have unclear expressions (for the pre-testing of face stimuli see [28]). Simply, their expressions are less salient. Thus, perceptual uncertainty of 'ambiguous' faces is not total but just greater than perceptual uncertainty of 'unambiguous' faces.

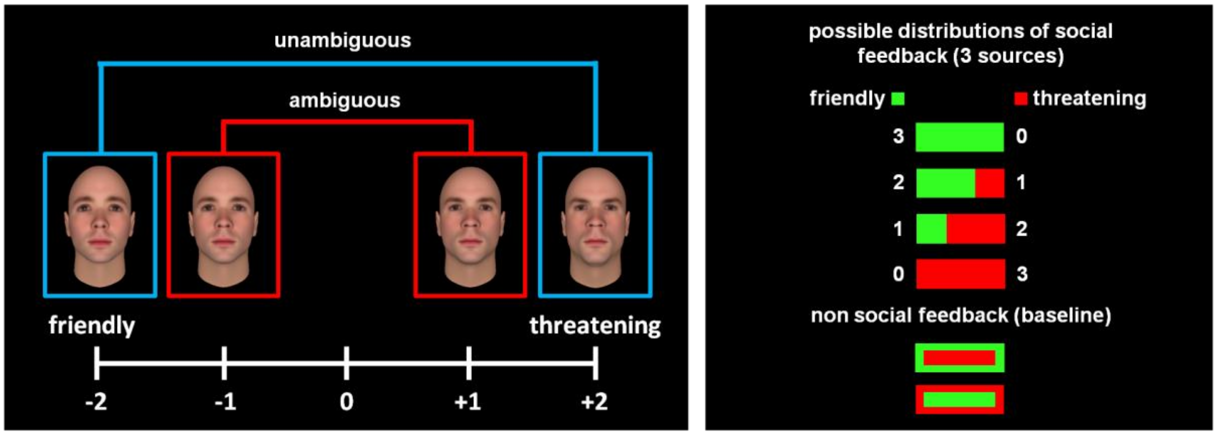


Figure 2. The left part of the figure represents an example of face stimuli. Faces located at the extremities of the continuum (-2; +2) are labelled 'unambiguous faces'. Those located around the center are labelled 'ambiguous faces'. The right part of the figure represents the possible distributions of social feedback. The fully coloured rectangles represent group consensus, meaning that the three influence sources made the same decision about the stimulus (green = 'friendly; red = 'threatening'). Intermediate feedback configurations are represented by rectangles coloured with ⅔ of one colour and and ⅓ of the other (2 'friendly' and 1 'threatening' or 2 'threatening' responses and 1 'friendly'). Finally, non social feedback in the baseline condition is represented by two superimposed green and red rectangles. These configurations do not provide any information about the decisions of the influence sources.

Manipulating the distribution of social feedback (see fig.2). Social feedback will be provided by three influence sources, represented in each trial by three personal photographs or three stills taken from video clips and located at the center of the second screen. The social feedback represents the decisions of the three sources on face presented at the beginning of the trial ('friendly' or 'threatening') and appears on the third screen in the form of a horizontal rectangle colored in red or/and green (green = 'friendly' decisions, red = 'threatening' decisions). Fully coloured rectangles represent group consensus, meaning that the three influence sources made the same decision. Intermediate feedback configurations are represented by rectangles coloured with ⅔ of one colour and and ⅓ of the other (2 'friendly' and 1 'threatening' or 2 'threatening' responses and 1 'friendly'). Group consensus and the intermediate feedback configurations will be randomized and fully counterbalanced across conditions. This will allow us to manipulate the conflict magnitude generated by social feedback (i.e., the effect of group consensus and disagreement on the participants' decisions). Baseline condition trials will consist of showing two superimposed green and red rectangles which model uninformative, non social feedback as no information about the influence sources' decisions can be inferred from them. Both the ambiguity of face stimuli and the meaning of social feedback will be pre-tested.

EEG acquisition. Several ERP components will be examined. The feedback error-related negativity (N_{FB}) and the feedback error-related positivity (P_{FB}) reflect the transmission of reinforcement signals from the mesencephalic dopamine system to the medial frontal cortex (i.e., rostral cingulate zone of the anterior cingulate cortex) [22]. The N_{FB} is involved in error monitoring and has been shown to be modulated by negative feedback following active-choices, especially when participants' expectations about the outcome of their choice conflict with the current feedback [29]. The P_{FB} has been shown to be modulated by negative feedback providing information relevant for the task goal [23]. These ERPs could be reliable predictors of a change in decisional strategy operating in medial frontal cortex after the processing of negative feedback. The N_{FB} and P_{FB} will by measured during the presentation of social feedback (see fig.1a,b,c).

We will also measure the early posterior negativity (EPN) and the late positive potentials (LPP). EPN and LPP have been shown to be modulated by the early and late processing of emotional stimuli such as friendly and threatening faces [24]. They have been localized in the centro-parietal and temporo-occipital regions, respectively. The EPN and LPP components will be examined during the second exposure to the face stimulus (decision phase) (see fig.1a,b). Task-related modulation of the EPN and LPP could predict that the processing of social feedback in the medial frontal zone exerts a backward influence on perceptual decisions operating in sensory and associative cortices.

Analyses. For both tasks 70 trials per conditions will be collected (2 stimulus levels [unambiguous vs. ambiguous] × 5 levels of feedback [baseline + 2 possible consensus + 2 possible intermediate distributions]) (see fig.2). We will compute hit and false alarm (FA) rates for each stimulus level – i.e., for both 'ambiguous' and 'unambiguous' faces. The hit rate is the proportion of 'friendly' responses when the stimulus was indeed friendly, or the proportion of 'threatening' responses when the stimulus was indeed threatening. The false alarm rate is the proportion of 'friendly' responses when the stimulus was threatening, or the proportion of 'threatening' responses when the stimulus was friendly. Hit and false alarm rates will be used to compute *d'* and *c* indices from Signal Detection Theory [30]. The *d'* is a measure of the subject's sensitivity (perceptual discrimination) and *c* is a measure of the subject's bias (decisional strategy).

A 2 × 5 repeated-measures ANOVA will then be performed on *d'* and *c*, with the within-subject factors 'Perceptual uncertainty' and 'Social feedback'. The confidence levels will be further analyzed using receiver operating characteristic (ROC) curves [31]. A sensitivity index to 'ambiguous' and 'unambiguous' faces will be determined using *A'*, the area under the curve. The group's mean amplitude of the N_{FB}, P_{FB}, EPN and LPP will be compared across each experimental condition.

Predictions. For the informational conformity task, the effect of social feedback should increase conformal decisions, as assessed by a decrease in *d'*. For example, participants should make fewer accurate decisions for threatening faces if those faces have been judged by the group as 'friendly'. This effect is expected to be restricted to faces eliciting perceptual uncertainty, and is expected to be independent of whether the feedback reaches consensus or not. Such a decrease in *d'* might be correlated with an increase in P_{FB} amplitude during the processing of social feedback. If so, it would suggest that in an informational context social feedback is not processed as conflicting information generating a prediction error signal, but rather as relevant information that participants can use to improve their performance – i.e., information that would be integrated into the decision process in order to reduce uncertainty. Finally, the decrease in *d'* might also be associated with an increase in EPN and LPP amplitudes recorded during the presentation of ambiguous faces after the feedback. This would reveal that, under condition of uncertainty, social feedback processed in medial frontal cortex generates expectations that bias perceptual decisions operating in sensory and associative cortices.

For the normative conformity task, a similar effect of social feedback on decisions is expected for faces that elicit perceptual uncertainty. However, social feedback should also have an effect on the decisions made on unambiguous faces, especially when associated with a consensus that conflicts with perceptual evidence. This change in decisional strategy should be revealed by a switch in the *c* index. For example, participants will likely categorize as 'threatening' unambiguous friendly faces that have been judged by the three sources as 'threatening'. We expect this switch in decisional strategy to be associated with an increase in the N_{FB} amplitude during the feedback delivery. If so, this would show that in a normative context social feedback that conflicts with perceptual evidence is processed in the medial frontal cortex as an erroneous action outcome or as a rewarding stimulus, both of which motivate behavioral adjustment. We do not expect to observed modulations of EPN and LPP amplitudes, suggesting that this change in decisional strategy does not affect lower level, perceptual decisional processes.

rTMS EXPERIMENT

COLLABORATION WITH THE INSERM UNIT 1028 – LYON, FRANCE

Objectives. In order to confirm the causal role of those brain structures examined in our social conformity tasks we will adapt the informational and normative conformity tasks to an on-line repetitive transcranial magnetic stimulation (rTMS) protocol.

Recently, it was shown that TMS-induced transient inhibition of the posterior medial frontal cortex (pMFC) attenuated social conformity [32]. However, nothing is known about whether pMFC disruption differentially affects conformity in an informational or normative context. Since we expect informational and normative contexts to modulate the effect of social feedback and perceptual uncertainty on decision-making our aim will be to test whether pMFC downregulation differentially affects behavioral performance in informational and normative contexts.

To do this we will collaborate with Dr. Alessandro Farnè and Dr. Karen Reilly at the Lyon Neuroscience Research Center (INSERM U1028 - CNRS UMR5292). This will give us the opportunity to interact with two acknowledged TMS specialists and to take use a state-of-the-art TMS platform.

Methods. A group of 20 healthy adults will be tested in both the informational and normative conformity tasks described above. During the tasks, we will use on-line, trial-by-trial rTMS [33] to transiently inhibit the right pMFC. On each TMS trial we will apply 5Hz rTMS over the target site for 1 sec after the first exposure to the face stimulus but before delivery of the social feedback (see fig.1). To control for any non-specific effects of rTMS, in addition to stimulating over the right pMFC we will also stimulate the right parietal cortex (precuneus). The stimulation sites will be chosen on the basis of anatomical MRIs acquired at the CERMEP platform in Lyon, as well as on the basis of brain coordinates used in a previous rTMS study [32]. Half of the trials will be with rTMS, the other half will be without TMS. Behavioral performances will thus be compared between trials with rTMS and trials without rTMS trials.

Predictions. We predict that transient inhibition of pMFC will reduce the number of conformal decisions. In particular, rTMS should affect decision-making by attenuating the perception of the conflict that a social feedback generates when it contradicts perceptual evidence. As such, the decrease in conformal decisions is expected to be particularly marked in the normative conformity task, in which participants are normally expected to follow the group decision even when it conflicts with perceptual evidence. By contrast, this effect should not be observed in the informational conformity task, since conformal decisions are motivated by the need to gain relevant information from social feedback in order to reduce perceptual uncertainty.

RELEVANCE OF THE PROJECT TO THE GOALS OF THE FOUNDATION

The results of the present project will provide the first insight into the cerebral bases of informational and normative conformity, and issues which is of particular importance to several scientific domains.

First, social conformity participates in the transmission of information between individuals and thereby significantly influences the evolution of human culture. Thus, exploring the neural dynamics of informational and normative conformity is of central importance for understanding what makes human cultural evolution possible. But culture also shapes brain, and it could be that some cultural traits have an influence on how individuals code and evaluate social information. A future development of this work would thus consist of testing whether brain mechanisms associated with the propensity to conform are sensitive to cultural variants (e.g. cultural variants that differently promote affective and intellectual autonomy or conservatism). Second, conformity regulates interactions between members of a group by generating shared expectations, and those who violate these expectations are exposed to social exclusion. A fine-grained understanding of brain mechanisms underlying conformity would contribute to better targeting of personality traits that are potentially predictive of anti-social and pro-social behaviors (e.g. psychopathic personality traits or sensitivity to interpersonal influences). For example, psychopathic patients exhibit dysregulation of the medial prefrontal cortex, a region that is known to be involved in the coding of social information as a rewarding stimulus [14-18]. At least partially, anti-social and pro-social behaviors could result from the incapacity to evaluate the value of social information. Testing the present tasks with populations exhibiting such personality traits is one of the promising potential extensions of this work.

This interdisciplinary project fits the goals of the Fyssen Foundation in several ways. First, at a theoretical level, it relies on various approaches to social conformity developed in neuroscience, social psychology, and cognitive anthropology. Second, at the experimental level, the project uses various methods originating in different fields such as experimental psychology and functional neuroimaging. Finally, I have been involved in interdisciplinary approaches to cognition since my undergraduate studies. I completed degrees in Philosophy, Cognitive Science and Clinical Neuropsychology, and I learned to test hypotheses using a wide range of methods. For these reasons I think that my project, as well as my profile, naturally fit Fyssen's interests.

REFERENCES

[1] Griffiths TL, Kemp C, and Tenenbaum JB (2008) Bayesian models of cognition. In Sun R (Ed.), The Cambridge handbook of computational psychology. Cambridge, UK: Cambridge University Press: 59-100.

[2] Chambon V, Domenech P, Pacherie E, Koechlin E, Baraduc P, Farrer C (2011) What are they up to? The role of sensory evidence and prior knowledge in action understanding. *PloS One* 6: e17133, 2011.

[3] Jacquet PO, Chambon V, Borghi AM, Tessari A (2012) Object Affordances Tune Observers' Prior Expectations About Tool-Use Behaviors. *PloS One* accepted.

[4] Jacquet PO, Tessari A, Binkofski F, and Borghi AM (2012) Can object affordances impact on human social learning of tool use? *Behav Brain Sci* in press.

[5] Acerbi A, Jacquet PO, Tennie C (2012) Behavioral constraints and the evolution of faithful social learning. *Current Zoology* in press.

[6] Sherif M (1935) A study of some social factors in perception. New York, NY: Columbia University Press.

[7] Asch SE (1955) Opinions and social pressure. Scientific American, 193(5): 31–35.

[8] Asch SE (1951) Effects of group pressure upon the modification and distortion of judgments. In: Guetzkow HS, ed. Groups, Leadership and Men Research in Human Relations. Pittsburgh: Carnegie Press, 177–190.

[9] Asch SE (1952) Social Psychology. New York: Prentice-Hall.

[10] Cialdini RB, Goldstein NJ (2004) Social influence: Compliance and conformity. *Ann Rev Psychol* 55: 591–621.

[11] Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. *Journal of Abnormal Social Psychology* 51(3): 629–636.

[12] Campbell JD, Fairey PJ (1989) Informational and normative routes to conformity: the effect of faction size as a function of norm extremity and attention to the stimulus. *Journal of Personality and Social Psycholgy* 57:457-468.

[13] Claidière N, Whiten A (2011) Integrating the Study of Conformity and Culture in Humans and Nonhuman Animals. *Psychol Bull* 138(1): 126-145.

[14] Izuma K, Saito DN, Sadato N (2008) Processing of Social and Monetary Rewards in the Human Striatum. *Neuron* 58:284–294.

[15] Izuma K, Saito DN, Sadato N (2009) Processing of the Incentive for Social Approval in the Ventral Striatum during Charitable Donation. *J Cog Neurosci* 22(4) : 621–631.

[16] Jones RM, Somerville LH, Li J, Ruberry EJ, Libby V, Glover G, et al. (2011) Behavioral and neural properties of social reinforcement learning. *J Neurosci* 31(37):13039-45.

[17] Campbell-Meiklejohn DK, Bach DR, Roepstorff A, Dolan RJ, Frith CD (2010) How the opinion of others affects our valuation of objects. *Curr Biol* 20:1165–1170.

[18] Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. *Neuron* 61(1):140-51.

[19] Berns GS, Chappelow J, Zink CF, Pagnoni G, Martin-Skurski ME, Richards J (2005) Neurobiological correlates of social conformity and independence during mental rotation. *Biol Psychiatry* 58(3):245-53.

[20] Rahnev D, Lau H, de Lange FP (2011) Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *J Neurosci* 31(29):10741-8.

[21] Summerfield C, Koechlin E (2008) A neural representation of prior information during perceptual inference. *Neuron*, 59(2):336-47.

[22] Nieuwenhuis S, Holroyd CB, Mol N, Coles MG (2004) Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci Biobehav Rev* 28(2):441-448.

[23] Band GPH, van Steenbergen H, Riddenrinkhof KR, Falkenstein M, Hommel B (2009) Action-effect negativity: irrelevant action effects are monitored like relevant feedback. *Biol Psychol* 82: 211-218.

[24] Schupp HT, Ohman A, Junghöfer M, Weike AI, Stockburger J, Hamm AO (2004) The facilitated processing of threatening faces: an ERP analysis. *Emotion* 4(2):189-200.

[25] Snyder M, Gangestad S (1986) On the nature of self-monitoring: matters of assessment, matters of validity. *Journal of Personality and Social Psychology* 51(1):125-39.

[26] Hare RD (1998) The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology* 3:101-122.

[27] Spielberger CD, Sydeman SJ, Owen AE, Marsh BJ (1999) Measuring anxiety and anger with the State-Trait Anxiety Inventory (STAI) and the State-Trait Anger Expression Inventory (STAXI). In M. E. Maruish (Ed.), The use of psychological testing for treatment planning and outcome assessment (2nd ed., pp. 993–1021). Mahwah, NJ: Lawrence Erlbaum Associates.

[28] Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. *Proc Natl Acad Sci USA* 105(32):11087-92.

[29] Kobza S, Thoma P, Daum I, Bellebaum C (2011) The feedback-related negativity is modulated by feedback probability in observational learning. *Behav Brain Res* 225: 396– 404.

[30] Macmillan NA, Creelman CD (2005) Detection theory: A user’s guide (2nd ed.). New York: Cambridge University Press.

[31] Swets JA (1973) The relative operating characteristic in psychology. *Science* 182: 990-1000.

[32] Klucharev V, Munneke MAM, Smidts A, Fernández G (2011) Downregulation of the posterior medial frontal cortex prevents social conformity. *J Neurosci* 31(33):11934-11940.

[33] Goebel SM, Calabria M, Farnè A, Rossetti Y (2006) Parietal rTMS distorts the mental number line: Simulating ‘spatial’ neglect in healthy subjects. *Neuropsychologia* 44:860-868.

APPENDIX III

List of publications and Curriculum Vitae

Manuscripts published or accepted in international peer-reviewed journals

Jacquet P.O., Chambon V., Borghi A.M., Tessari A. (in press). Object affordances tune observers' prior expectations about tool-use behaviors. *PLoS One*

Jacquet P.O., Tessari A., Binkofski F., Borghi AM. (in press). Can object affordances impact on human social learning of tool use? Comment article. *Behavioral and Brain Sciences*

Chambon V., Domenech P., Barbalat G., Pacherie E., **Jacquet P.**, Farrer C. (in press). Reply: the Bayesian equation. *Brain*.

Acerbi A., **Jacquet P.O.**, Tennie C. (2012). Behavioral constraints and the evolution of faithful social learning. *Current Zoology, 58*(02), 307-318.

Sablier J., Stip E., **Jacquet P.**, Mobus Group, Franck, N. (2012). Ecological Assessments of Activities of Daily Living and Personal Experiences with Mobus, An Assistive Technology for Cognition: a Pilot Study in Schizophrenia. *Assistive Technology*, *24*, 1-11.

Chambon V., Barbalat G., Pacherie E., **Jacquet P.**, Koechlin E., Franck N., Farrer C. (2011). Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia. *Brain, 134(12)*, 3728-41.

Manuscripts published in French-speaking peer-reviewed journals

Jacquet P.O., Franck N., Demily C. (2012) Déficits du monitoring de la source et hallucinations schizophréniques. *EMConsulte,* 37-091-A-20.

Demily C., **Jacquet P.**, Marie-Cardine M. (2009). How to differentiate schizophrenia from bipolar disorder using cognitive assessment? *L'encéphale*, *35(2)*, 139-45.

Manuscript under revision in international peer-reviewed journals

Scorolli C., **Jacquet P.O.**, Binkofski F., Nicoletti R., Tessari A., Borghi A.M. Abstract and Concrete Sentences Processing Differentially Modulates Corticospinal Excitability

Jacquet P.O., Avenanti, A. Perturbing the Action Observation Network during perception and categorization of others' actions: state-dependency and virtual lesion TMS effects.

Jacquet P.O., Roy, A.C., Chambon, V., Borghi, A.M., Farne, A., Reilly, K.T. Modulating human motor resonance: Exposure to suboptimal actions suppresses corticospinal excitability.

Jacquet P.O., Chambon V., Franck N., Demily C. 'I didn't say it, but I thought it aloud': Internal source-monitoring ambiguity and its relationship with formal thought disorders of schizophrenic patients.

Pierre O. JACQUET

PHD student – Neuropsychologist

72, rue de Sèvres
75007 PARIS
Telephone: +33(0)650493008
e-mail: pierre.jacquet3@unibo.it

PERSONAL DETAILS

Gender: Male
Date of birth: 28 February 1980
Citizenship: French

RESEARCH INTERESTS

Motor and social cognition; Cultural evolution; Psychopathology

EDUCATION

Bologna University, Italy

2009-present: Joint International Cognitive Neuroscience PHD Program

Project title: *Social learning and action understanding in human observers: contributions of sensori-motor constraints and prior information.*

Mentor: Pr. Anna M. Borghi

Strasbourg I University, France

2006-2007: Master degree in Clinical and Cognitive Neuropsychology, graduated with distinction.

Lyon II University, France

2005-2007: Master degree in Cognitive Sciences. Specialization: Research in Neuropsychology, graduated with distinction.

Dissertation title: *Role of encoding source on the production of familiarity and recollection judgements: A study in patients with schizophrenia.*

Mentor: Pr. Nicolas Franck.

2002-2004: BA in Psychology, graduated with distinction.

Grenoble II University, France

1999-2001: BA in Philosophy.

RESEARCH EXPERIENCES

« ImpAct » – INSERM, U1028 – Lyon, France

September, 2010-present: Visitor PhD student at the INSERM U864 of Lyon, France.

Mentor: Dr. Alessandro Farnè, Dr. Karen T. Reilly

Dipartimento di Psicologia – Università di Bologna, Bologna, Italia

May, 2008-June, 2011: Research Assistant/PhD student. *European project ROSSI - Emergence of Communication in RObots through Sensorimotor and Social Interaction.*

Mentor: Pr. Anna M. Borghi.

Institut des Sciences Cognitives – Centre de Neuroscience Cognitive, CNRS UMR5229 – Lyon, France

September, 2007-April, 2008: Research Assistant. *Evaluation of the impact of cognitive orthotic on mnesic and executive functionning of schizophrenic patients.*

Mentors: Pr. Nicolas Franck ; Pr. Emmanuel Stip (Montréal University, Canada).

September, 2007-present: Associate investigator on the *RECOS project - COgnitive REmediation program adapted for Schizophrenic patients.*

Mentors: Pr. Nicolas Franck. ; Pr. Hélène Verdoux (Bordeaux II University, INSERM U657)

WORKING EXPERIENCES IN CLINICAL NEUROPSYCHOLOGY

Hôpitaux Universitaire de Genève, Genève, Switzerland
June, 2005 - September, 2006: Neuropsychologist trainee. Unit of dementia’s detection.

Centre Hospitalier le Vinatier, Lyon, France
September, 2007 – April 2008: Neuropsychologist. Adult Psychiatry Unit.

FOREIGN LANGUAGES

French: native
English: fluent
Italian: fluent

EXPERIMENTAL TECHNICS

During these last 4 years, the different projects I was involded in allowed me to master different skills such as:

TMS stimulation using Magstim 200 and Magstim Rapid^2 stimulators.

Surface electrodes recording using SPIKE 2/CED and Biopac Student Lab MP36 data collection interfaces.

Stereotaxic neuronavigation system interfaced with the SoftTaxic (EMS) software.

COMPUTER AND SOFTWARE SKILLS

Statistical Analysis : Statistica

Running and Preparing Experiments : Presentation, E-Prime2

Manipulating Videos, Images and Sounds : Adobe Premiere Pro 2.0 ; Reason 3.0

GRANTS

2010 – Marco Polo doctoral mobility fellowship of the University of Bologna (3500€)

2008 – Hospital Program of Clinical Research (PHRC). Scientific collaborator and co-writer of the RECOS project – COgnitive REmediation program adapted for Schizophrenic patients (220k€).