What Do Mirror Neurons Contribute to Human Social Cognition?

PIERRE JACOB

Abstract: According to an influential view, one function of mirror neurons (MNs), first discovered in the brain of monkeys, is to underlie third-person mindreading. This view relies on two assumptions: the activity of MNs in an observer's brain matches (simulates or resonates with) that of MNs in an agent's brain and this resonance process retrodictively generates a representation of the agent's intention from a perception of her movement. In this paper, I criticize both assumptions and I argue instead that the activity of MNs in an observer's brain is enhanced by a prior representation of the agent's intention and that their task is to predictively compute the best motor command suitable to satisfy the agent's intention.

Introduction

One of the most remarkable contributions of cognitive neuroscience in the past fifteen years or so has been the discovery of so-called 'mirror neurons' (MNs) in the ventral premotor cortex of macaque monkeys and the subsequent discovery of a 'mirror system' in the human brain.[1] MNs are sensorimotor neurons that fire both when an animal is executing some kinds of hand or mouth action directed

---

Address for correspondence: Institut Jean Nicod, Ecole Normale Supérieure, Pavillon Jardin, 29, rue d'Ulm, 75005 Paris, France.
Email: jacob@ehess.fr

---

[1]   Ramachandran (2000) has gone as far as claiming that 'mirror neurons will do for psychology what DNA did for biology … provide a unifying framework and help explain a host of mental abilities that have hitherto remained mysterious and inaccessible to experiments'. The obvious question is: which mental abilities will be accounted for by MNs?

towards a target and when the animal is observing a conspecific (or an experimenter) perform the same kind of actions. Their activity has been alleged to underlie such cognitive capacities as imitation learning (Rizzolatti *et al.*, 2001; Rizzolatti and Craighero, 2004), mindreading (Gallese and Goldman, 1998), and even language understanding (Rizzolatti and Arbib, 1998). Here, I will concentrate on mindreading.

Roughly speaking, the reasoning that led Gallese and Goldman (1998) to the view that MNs might underlie mindreading involves the following three steps. First, the activity of MNs is treated as a replicative or resonance mechanism in which the firing of MNs in an observer's brain resonates with (or matches) the firing of MNs in the agent's brain.[2] Secondly, by duplicating the activity of MNs in an agent's brain, the activity of MNs in an observer's brain is taken to constitute an automatic mental simulation of the agent's observed movements: the activity of MNs in the observer's brain enables the observer to match the agent's observed movements onto her own motor repertoire (without executing the movements in question). Finally, by performing a mental simulation of the agent's observed movements, the activity of MNs is seen as enabling the observer to recognize the agent's action, or even to represent her intention (or goal). Since representing an agent's intention is unquestionably part of third-person mindreading, it turns out that one fundamental function of MNs is to underlie mindreading.

As the above reconstruction makes clear, the reasoning used to link MNs to mindreading borrows concepts from the simulation approach to mindreading. Conversely, it might be—and has been—argued that the discovery of MNs vindicates the simulation approach to mindreading by showing that there are neural simulation mechanisms in the primate brain which enable an observer to make sense of an agent's action based on the perception of her movements.

I agree that if the activity of MNs in an observer's brain did generate a representation of an agent's intention, then MNs would contribute to mindreading. However, I think that two of the steps used to reach this conclusion are questionable. I will assume for the sake of argument that by replicating the activity of MNs in an agent's brain, the activity of MNs in an observer's brain constitutes a mental simulation of the agent's observed movements. But, first, it is highly questionable whether by mentally rehearsing an agent's observed movements, an observer could represent the agent's underlying intention. Secondly, some recent evidence casts doubt on the assumption that the activity of MNs is a replicative (or resonance) process and suggests instead that it is a predictive process. Since I argue against the view that the activity of MNs underlies mindreading by virtue of representing an agent's intention, I must provide an alternative answer to the question: what is the function of MNs? Representing an agent's action involves at least two complementary parts: representing the agent's intention and representing the

---

[2] I will use the verbs 'to resonate', 'to match', 'to rehearse', 'to replicate' and 'to duplicate' interchangeably.

motor command by which this intention is achieved.[3] I argue that the activity of MNs contributes to the latter, not to the former. Furthermore, whereas an agent's intention (which is the cause of her motor act) can only be retrodicted from the perception of her motor act, the motor command that will generate her next move can be predicted. Thus, I argue that the activity of MNs is predictive, not retrodictive.

In the first section, I shall review the evidence adduced to support the view that what I shall call 'interpersonal mirroring' is a replicative process, or that the activity of MNs in an observer's brain matches the activity of MNs in the agent's brain. In the second section, I shall examine the simulation approach to mindreading generated by incorporating MNs as processes of neural simulation. In the third section, I shall argue that motor resonance is neither a necessary nor a sufficient mechanism for representing an agent's intentions. I will further show that very recent evidence has led MN theorists to endorse a modified view of MNs, on which the function of interpersonal mirroring turns out *not* to be replicative but rather predictive. If MNs are predictive, then what they predict must be the agent's next move, not the agent's intention that caused her observed movement. In the last section, I shall argue for an alternative view (already considered by Csibra, 2005), on which the function of MNs is to compute (and thereby predict), not the agent's intention, but the motor command suitable for achieving the intention.

1. The Direct Matching or Resonance Model of Action Understanding

As noted above, MNs were first discovered in the ventral premotor cortex of macaque monkeys. The reason that MNs were so-called is that they are sensorimotor neurons with both motor and perceptual properties: they fire both when an animal is executing some kinds of transitive (i.e. directed towards a target) hand and/or mouth action and when the animal is observing a conspecific (or a human experimenter) perform the same kind of action. The perceptual properties of MNs are said to 'mirror' their motor properties.[4] As a result, when an animal observes a conspecific perform a transitive action, the activity of MNs in her brain is said to 'mirror' the activity of MNs in the agent's brain. This interpersonal mirroring appears to be a matching of brain activity, enabling an observer to perform a motor simulation of the agent's observed movements without executing them. In any case, MNs deserve to be so-called only if the congruence between their perceptual and motor properties is statistically strong enough—a point disputed by Csibra (2004).

---

[3] As I will show in section 4.3, cognitive neuroscience suggests that the intuitive notion of having a goal can be decomposed into a motor and a sensory representation.
[4] As noticed by Sperber (2004), the expression 'mirror neurons' has understandably caught people's attention and triggered their imagination. For mouth actions, see Ferrari *et al.*, 2003.

1.1 MNs in Monkeys

The ventral premotor cortex of macaque monkeys contains not only MNs but also so-called 'canonical' neurons, whose motor properties are the same as those of MNs (they fire when the animal executes transitive hand and mouth actions), but which differ in their perceptual properties. Unlike MNs, canonical neurons do not fire when the animal perceives an action directed towards a target; rather, they fire when the animal perceives an object that can be the target of an action. There are also cells in the Superior Temporal Sulcus (STS) which have perceptual properties in common with MNs (they also fire in response to the perception of actions performed by conspecifics), but lack corresponding motor properties: they only fire when the recorded animal is observing another agent, never when he performs an action himself (Perrett *et al.*, 1982; 1989; Keysers and Perrett, 2004).[5]

For a decade, MN theorists have reported experiments revealing the congruence between the motor and perceptual properties of MNs. Umiltà *et al.* (2001) have shown that MNs in the ventral premotor cortex fire both when the animal executes a transitive hand action and when he observes such a hand action performed by another and knows that there is a target, even if the target is hidden behind an occluder. However, if there is no target and the animal knows it, then MNs fail to fire. (On the cross-modal properties of MNs in perceptual tasks, see Kohler *et al.*, 2002; Keysers *et al.*, 2003.) Thus, the evidence shows that in monkeys, the activity of MNs in an observer's brain is restricted to the perception of actions directed towards a physical target.

1.2 The Mirror System in Humans

Since the discovery of MNs in the monkey brain, much evidence from brain imaging has revealed the existence of a 'mirror system' network in humans (cf. Rizzolatti *et al.*, 2001 and Rizzolatti and Craighero, 2004). This evidence suggests that the activity of the mirror network in humans is involved in the execution and observation of a wider class of actions than in non-human primates, including intransitive actions (not directed towards a target) such as pantomimes and ostensive communicative gestures. Of the numerous experiments, I will merely mention three of particular note.

In a TMS (transcranial magnetic stimulation) study, Fadiga *et al.* (1995) have shown that the observation of hand actions executed by others prompted a motor facilitation in the very same arm and hand muscles that were being used by the agents of the observed actions.[6] In an fMRI study, Buccino *et al.* (2001) showed that, compared to observation of a static face, hand or foot, observation of both transitive and intransitive actions involving movements of the hand, mouth or foot led to the somatotopic activation of the premotor cortex. In another fMRI study,

---

[5] In fact, they respond to the perception of a wider class of actions than MNs.
[6] However, Romani *et al.* (2005) also report motor facilitation prompted by the observation of *biologically impossible* finger movements.

Buccino *et al.* (2001) had subjects watch silent films showing a human, a monkey or a dog perform either a transitive or an intransitive action (the transitive action was biting food, the human intransitive action was silent speech, and the dog's intransitive action was barking). They found, on the one hand, that observation of a transitive action by any agent led to global activation in the areas involved in the mirror system, and on the other hand, that the activation of motor areas increased as a function of whether the observed action could be matched onto the observer's motor repertoire: it was significantly stronger for human silent speech than for dog barking.

1.3 MNs and Action Understanding

Based on this evidence, MN theorists have argued for a basic distinction between two ways an observed action can be represented and understood (cf. Rizzolatti *et al.*, 2001). Actions involving movements that belong to the observer's motor repertoire can be mapped onto the observer's motor system. Observation of such actions causes the observer's motor system to *resonate*. By mapping an agent's observed movements onto her own motor repertoire (i.e. by motor resonance), an observer achieves a distinctive kind of 'engaged' or immediate understanding. This is the *resonance* model of action understanding.

By contrast, actions that do not belong to the observer's motor repertoire and cannot be so mapped must be categorized on the basis of their visual properties. For example, a primate may have the capacity for a detached visual analysis of a bird's flight, but lack a motor understanding of it, since a primate cannot match a bird's wing movements onto his own motor repertoire. Actions that cannot be understood by motor resonance (or MN activity) will be recognized through the activity of purely perceptual brain areas (such as the inferotemporal lobe and the STS). The understanding thereby achieved lacks the immediacy of the understanding delivered by motor resonance. As Gallese *et al.* (2004) write:

> We will posit that, in our brain, there are neural mechanisms (mirror mechanisms) that allow us to directly understand the meaning of the actions and emotions of others by internally replicating ('simulating') them without any explicit reflective mediation [...] The fundamental mechanism that allows us a direct experiential grasp of the mind of others is [...] direct simulation of the observed events through the mirror mechanisms [...] A crucial element of social cognition is the brain's capacity to directly link the first- and third-person experiences of these phenomena (i.e. link 'I do and I feel' with 'he does and he feels'). We will define this mechanism simulation (Gallese *et al.*, 2004, p. 396).[7]

---

[7] As this makes clear, Gallese *et al.* (2004) extend the role of mirroring processes to an immediate understanding of others' emotions. For reasons of space, I shall not examine this interesting extension here, which is used as a major step by Gallese (2003, 2004) and Goldman (2004) in supporting the view that not all mirroring processes are motor processes, but see section 2.4.

As the quote makes clear, Gallese *et al.* (2004) assume that the activity of MNs in an observer's brain constitutes an internal replication of the agent's observed movements and that such internal replication yields an engaged form of understanding of the observed action, i.e. an understanding that reflects the fact that the observer has the motor resources for producing tokens of the type of bodily movements performed by the agent. By contrast, a purely visual analysis of the components of the action does not deliver such an engaged (motor) understanding. However, it is one thing to distinguish between an engaged and a disengaged (or detached) understanding of an agent's action, based on motor resonance, and quite another thing to make the further and stronger claim that motor resonance also yields a 'direct experiential grasp of the [agent's] mind'. This latter claim is stronger because, unless one subscribes to behaviorism, it does not seem as if being able to match an observed movement onto one's motor repertoire is sufficient for knowing what the agent has in mind.[8] So Gallese *et al.*'s (2004) stronger claim raises the following question: could the engaged understanding of an agent's action based on motor resonance constitute an instance of third-person mindreading?

2. The Scope of the Mental Simulation Approach to Mindreading

What Baron-Cohen (1995) has called 'mindreading' is the cognitive ability to represent the psychological states (perceptions, emotions, intentions, desires, beliefs, etc.) of oneself and others.[9] Most philosophers and psychologists assume that healthy human adults make extensive social use of their mindreading ability in describing, explaining and predicting their own and others' actions. At this stage, the exact nature of the cognitive mechanisms underlying human mindreading remains controversial, and so do their phylogenetic history and ontogenetic development. Nonetheless, one thing seems clear: to perform a third-person mindreading task is to form a belief about another's psychological state. So the question arises: could what Gallese *et al.* (2004) call a direct experiential (non-conceptually mediated) grasp of another's mind be classified as an instance of third-person mindreading?

In an insightful discussion, Goldman (2004, 2006, pp. 133-37) has expressed doubts. On Goldman's view, mirroring might be necessary for reading another's mind, but cannot be sufficient.[10] Mirroring (or motor resonance) is the phenomenon

---

[8] This point has recently been made by Borg (2007).

[9] Although full-blown human mindreading can be both third-personal and first-personal, in this paper, I deal exclusively with third-person mindreading. See e.g. Nichols and Stich, 2003 and Goldman, 2006.

[10] As I will explain in this section, Goldman's (2006) view that mirroring is a necessary condition for mindreading follows from his views (i) that mirroring is one instance of mental simulation and (ii) that mental simulation is a basis for mindreading. In section 3.1, I shall question the assumption that mirroring is *necessary* for mindreading.

whereby an observer's brain activity is caused by similar brain activity in the agent: the activity in the agent's brain causes her to make a movement, and the observer's perception of this movement causes the observer to undergo an analogous brain activity. (I take it that an observer's premotor cortex could not resonate with another's action if the observer could not perceive the agent's motor act.)[11] According to Goldman, mirroring (or motor resonance) is not sufficient for mindreading because mindreading requires attribution of some psychological state or other to the agent (who is being mindread). If performing a third-person mindreading task involves forming a belief about another's psychological state, then unless she possessed the concept of the relevant psychological state, the mindreader could not impute it to the agent.

In the 1980's, the empirical investigation of human mindreading provoked a lively debate between two main approaches: the 'theory-theory' (TT) and the 'simulation' (ST) approaches. According to TT approaches, mindreading is a kind of detached naïve psychological theorizing based on tacit knowledge of folk psychological laws relating unobservable psychological states to one another, to observable (sensory) inputs and to observable (behavioral) outputs.[12] By contrast, ST approaches assume that what underlies mindreading is psychological similarity. Central to ST accounts are the concept of *pretence* and the thesis that what enables an individual to engage in third-person mindreading tasks is not her tacit knowledge of naive psychological laws, but the fact that she shares the very same cognitive resources as the individual whom she is trying to mindread. According to ST accounts, third-person mindreading tasks involve use of these shared cognitive resources 'off-line' for purposes of imaginative pretence.[13] An advocate of ST might thus expect there to be brain mechanisms that generate a representation and understanding of an agent's intention by mentally rehearsing her observed action.

So the discovery of MNs in the 1990's raised the intriguing prospect that what I earlier called 'interpersonal mirroring' processes might be novel instances of mental simulation, in addition to pretence. This would broaden the explanatory

---

[11]   An anonymous referee interestingly suggests that an observer (e.g. a reader of a hand-written text) might be said to 'resonate' with an agent's act (e.g. the author's hand-writing) if she merely imagined (rather than perceives) the agent's act. But if so, then this would be resonance in a loose sense. It is not clear at all that resonance in this sense is available to non–human primates, in whose brains MNs were discovered and who cannot either write or read.

[12]   For purposes of the present paper, which is entirely devoted to extensions of ST accounts of mindreading, TT accounts are adequately characterized as psychological versions of the functionalist view of psychological states, earlier developed by philosophers of mind. Cf. Gopnik and Meltzoff, 1997; Gopnik and Wellman, 1992, 1994.

[13]   The original emphasis of ST accounts on pretence is reminiscent of Quine's (1960) view that indirect quotation involves an 'essentially dramatic act' in which we project ourselves into the speaker's state of mind. ST's emphasis on similarity of cognitive resources as the basis of third-person mindreading is reminiscent of the appeal to *Verstehen* (a non-causal understanding of an agent's reasons) and *Einfühlung* (empathetic understanding) by late 19th century German philosophers and psychologists, some of whom endorsed the methodological dualism between the natural sciences and the humanities. Cf. Currie, 1998; Goldman, 1989, 1992, 2006; Gordon, 1986, 1992; Heal, 1986. For informative discussion, see Nichols and Stich, 2003.

scope of ST approaches to human mindreading and strengthen its empirical basis; it would also suggest that a fairly elementary mechanism in the motor cognition of non-human primates might constitute the neural basis or a phylogenetic precursor of human mindreading. In 1998, Vittorio Gallese (one of the cognitive neuroscientists who discovered MNs) and Alvin Goldman (one of the philosophical advocates of ST approaches to mindreading) published an influential paper that linked the activity of MNs to mindreading for the first time by adopting an ST approach to mindreading and arguing that MNs are instances of neural simulation. Gallese and Goldman's (1998, p. 498) conjecture was that 'MNs represent a primitive version, or possibly a precursor in phylogeny, of a simulation heuristic that might underlie mindreading'. On this view, MNs would be a neural basis and a phylogenetic precursor of mindreading because the former is a resonance mechanism and the latter is enhanced by neural similarity.

MNs were first discovered in monkeys. The evidence does not unequivocally show that monkeys engage in mindreading tasks.[14] Could MNs constitute a primitive version of a mental simulation heuristic, as Gallese and Goldman conjecture? Since most early ST approaches to mindreading appealed to the concept of pretence, one obvious challenge for Gallese and Goldman's ambitious research program is to show that both pretence and interpersonal mirroring, exemplified by MNs, are instances of mental simulation. This is the challenge to which I turn in the present section. There are two different strategies for implementing Gallese and Goldman's (1998) research program and, as I shall shortly argue, they part company on how exactly to fill the details of this program.

2.1 Two Meanings of 'Simulation'

At the core of ST accounts of mindreading is the view that pretence is an instance of mindreading. Gallese and Goldman (1998) were the first to propose that mirroring processes may be novel instances of mental simulation. However, the word 'simulation' has been used in two different scientific contexts with different meanings. Thus, the notion of mental simulation could be broadened in two slightly different ways, emphasizing one or other of the two meanings.[15]

The word 'simulation' is widely used in computer science to mean, roughly, 'modeling' and/or 'computation'. This sense of 'simulation' is the one recently used by computational theorists of action, who have developed so-called 'internal models' of action (cf. Blakemore *et al.*, 2000; Jeannerod, 2001; Miall, 2003; Wolpert, 1997; Wolpert and Gharamani, 2000; Wolpert *et al.*, 2001). The English word 'simulation' also derives from the Latin verb *simulare*, 'to imitate' (or 'to

---

[14] Povinelli and Vonk, 2004; Tomasello and Call, 1997.
[15] The richness (or looseness) of the meaning of 'simulation' has led Nichols and Stich (2003, p. 134) to conclude that 'the term has become quite useless. It picks out no natural or theoretically interesting category'. For a response, see Goldman, 2006, p. 35.

duplicate'), and the Latin adjective *similis*, 'similar', and this gives rise to a second, 'similarity' sense.[16] Since the crucial assumption of pretence-based ST approaches has been that third–person mindreading is based on psychological similarity between the mindreader and her psychological target, the most relevant meaning of 'simulation' for pretence-based ST approaches to mindreading would seem to be the *similarity* meaning.[17]

However, Goldman (2006, pp. 35–39) points out that there is an important difference between the notion of simulation and such notions as similarity and duplication: whereas the latter are symmetrical, the former is not. Following Goldman (2006, p. 37), I shall assume that process *P* can only simulate process *P'* if *P* has the function of duplicating *P'* (in the relevant respects).[18]

2.2 Pretence

Pretence-theoretic accounts of third–person mindreading satisfy the ST assumption that third–person mindreading is an attempted duplication of another's psychological life based on similarity between their relevant cognitive processes. A mindreader can use her shared cognitive resources 'off–line' for purposes of imaginative pretence.[19] For example, if the task is to predict another's decision, then on ST accounts, a mindreader creates in her own mind so-called 'pretend' beliefs and desires, whose contents are expected to be relevantly similar to those of the person whose decision she is trying to predict. She then feeds these pretend beliefs and desires as inputs to her own decision–making mechanism, and by running this mechanism 'off–line', is able to mentally represent a pretend decision that is likely to be relevantly similar to the other person's. Of course, she does not act on the basis of this representation, but it enables her to form a belief about the other person's decision.[20]

According to Goldman (2006), mental simulation may be either intrapersonal (involving simulation of one's own psychological state) or interpersonal (involving simulation of another's psychological state). Pretence-theoretic accounts apply to both intrapersonal and interpersonal mental simulation. Third-person mindreading tasks would qualify as interpersonal mental simulation, while visual and motor imagery tasks would be instances of intrapersonal mental simulation.[21] There is

---

[16]   Cf. Goldman, 2006, pp. 35–36.
[17]   In her seminal paper, Heal (1986) used 'replication' for 'simulation' and Goldman (2006, ch. 2) is very explicit on this point.
[18]   Any appeal to similarity (or resemblance) must specify which are the *relevant* respects. But I shall not press ST theorists of mindreading on this point here.
[19]   Cf. Currie, 1998; Goldman, 1986, 1992; Gordon, 1986; Heal, 1986.
[20]   Cf. Stich and Nichols, 1992.
[21]   Cf. Currie (1995) and Currie and Ravenscroft (1997) on mental visual imagery and mental motor imagery, respectively. To my knowledge, no advocate of pretence-based accounts of mindreading suggests that intrapersonal mental simulation sub-serves first-person mindreading tasks (i.e. results in beliefs about one's own psychological properties or psychological self-knowledge).

evidence that the primary visual system is activated in mental visual imagery tasks (where one imagines seeing something). If so, then such tasks would involve the intrapersonal creation of a neural similarity with some of the brain activities underlying visual perception. According to pretence-theoretic accounts, in mental visual imagery, the visual system is used off-line: it takes pretend inputs from memory (as opposed to retinal inputs) and produces as a pretend output a visual image (as opposed to a visual percept).[22]

2.3 Differences Between Mirroring and Pretence

From the first, the standard view among MN theorists has been that, although an observer does not (or need not) execute the observed action, the activity of MNs in her brain matches the activity of MNs in the agent's brain. MN theorists have emphasized the *congruence* between the motor and perceptual properties of MNs (or the replicative character of interpersonal mirroring). If the function of interpersonal mirroring is indeed to replicate brain activity, then interpersonal mirroring would qualify as another instance of mental simulation on the ground that its function is to generate similarity of brain activity across two individuals.

However, as noticed by Gallagher (2007) and recognized by Goldman (2006), there are at least two relevant differences between mirroring as exemplified by MNs and either intrapersonal or interpersonal pretence. First, whereas the former is automatic, involuntary and non-conscious, most instances of both intrapersonal and interpersonal pretence are highly controlled, voluntary and conscious. Secondly, successful pretence involves the endogenous creation of a resemblance relation between two mental states (of the same mind or two different minds). By contrast, in interpersonal mirroring processes, the activity of one individual's brain is exogenously caused by the activity of another individual's brain. The interpersonal resemblance (or resonance) between brain activities achieved by interpersonal mirroring is thus reminiscent of a process of motor contagion.[23]

To sum up, the discovery of MNs creates a challenge for ST accounts of mindreading. It offers the opportunity to extend the class of mechanisms potentially relevant to the ability to engage in third-person mindreading tasks. But then ST accounts must accommodate the significant differences between pretence and

---

[22] In motor imagery tasks, one imagines performing an action without overtly executing it. There is much evidence that the primary motor system is active in such tasks. Cf. Jeannerod, 1994, 1997; and Pacherie, 2000, among many sources. Interestingly, Sirigu and Duhamel (2001) found dissociations in both healthy subjects and brain-damaged patients between visual imagery (or so-called 'third-person' tasks) and motor imagery (or so-called 'first-person' tasks). Cf. section 4.3.

[23] Motor contagion is the causal process whereby the perception of another's behavior automatically causes the perceiver to exemplify the perceived behavior. For example, the process whereby infants are caused to cry by perceiving other infants' cries is an instance of motor contagion. So is yawning in human adults, cf. Provine, 1989. But one question that arises is: at which level in the representation of a complex observed behavior does motor contagion apply? See the end of section 3.2 for some discussion.

mirroring within a single generic concept of mental simulation without losing the insights of pretence-theoretic accounts of mindreading. I shall consider two available strategies for doing this.

2.4 'Embodied Simulation'

One strategy, suggested by Gallese (2003, 2004, 2005), Gallese and Metzinger (2003) and Gallese and Lakoff (2005), involves embedding the notion of mental simulation within the fast growing research program of so-called 'embodied cognition'. Advocates of embodied cognition reject the classical computational-representational view of human cognition, with its view of the human mind as a 'disembodied' Turing machine or a syntactic engine manipulating symbols according to syntactic rules. What Gallese (2003, 2004, 2005), Gallese and Lakoff (2005) and Gallese and Metzinger (2003) mean by 'embodied simulation' is an 'automatic, unconscious and pre-reflexive' process with very wide potential application, which is seen as underlying our 'general capacity of predicting upcoming sensory events' and our mastery of concepts. Embodied simulation is described as 'a basic functional mechanism of our brain', which plays 'a major role in our epistemic approach to the world' by enabling us 'to model reality'; its function is 'the modeling of objects, agents and events', i.e. 'modeling the interactions between a situated organism and its environment'.

In particular, Gallese and Lakoff's (2005, p. 3) explicit purpose is to provide a 'testable embodied theory of concepts … capable of reconciling both concrete and abstract concepts within a unified framework'. Embodied simulation is thus seen as encompassing, but going well beyond, 'our understanding of interpersonal relations'. For example, Gallese and Lakoff (2005, p. 15) surmise that the activity of 'canonical neurons' (described in section 1.1 above) could 'underpin basic-level categories of objects' by 'bringing together' their 'perceptual and motor properties'. An obvious challenge for this view is to determine what could be the 'motor programmes that define the prototypical interaction' with instances of the basic-level concepts such as CAT, TREE or STAR.[24]

Thus, one important question for the 'embodied simulation' strategy is: to what extent does embodied simulation involve motor processes? Gallagher (2007), Jacob and Jeannerod (2005) and Mahon and Caramazza (2005) argue that it does. For example, Mahon and Caramazza interpret the embodied simulation strategy (defended by Gallese and Lakoff, 2005) as involving the two-part claim that motor production processes underlie both the recognition of visually presented actions and the representation of conceptual knowledge of objects and actions. They provide neuropsychological evidence against the first part of this claim, based on double dissociations between human patients with apraxia but without pantomime

---

[24] Bloom (1996) also offers convincing evidence against the claim that the contents of artifact concepts (e.g. CHAIR) could arise from the 'convergence of gestalt object perception … and motor programmes that define the prototypical interaction with the object'.

agnosia—who cannot use tools correctly, but can recognize actions and/or pantomimes involving use of tools—and those with pantomime agnosia but without apraxia, for whom the reverse is true. They also argue that a double dissociation between optic ataxic patients (who are impaired in visually guided actions of prehension, but whose visual recognitional abilities are preserved) and visual form agnosic patients (for whom the reverse is true) is evidence against the second part of the claim, that motor production processes underlie the visual recognition of objects.[25]

Gallagher (2007), who accepts a so-called 'enactivist' view of the perception of human actions while rejecting a simulation-based account of mindreading, seems committed to the claim that all mindreading involves some motor processes. According to enactivist accounts, to perceive an object is to act upon it, to know how to act on it, or to know the sensory consequences of one's own actions towards it (cf. O'Regan and Noë, 2001 and Noë, 2004).[26] Unlike the visual perception of an inanimate graspable object, however, the perception of a human action causes a neural process of mirroring.[27] Now, on Gallagher's view, mirroring (or 'neural resonance') processes triggered by the perception of another's action are straightforward perceptual processes, but they do not involve simulation. On the one hand, he accepts a perceptual model of interpersonal understanding on which an agent's psychological state (e.g. his intention) can be directly perceived: by perceiving a human agent act, we directly perceive his intention. On the other hand, such a perceptual process is a motor process because, on the enactive account, the perception of a human action—like the perception of anything else—is a motor process.[28]

All this raises a problem for Gallese's (2003, 2004, 2005) version of embodied simulation. On the one hand, Gallese (2005) and Gallese *et al.* (1999) have expressed explicit reservations towards the two-visual systems model of primate vision, according to which there is a dichotomy between the visual perception of graspable objects and visually guided actions directed towards these objects. As Gallese *et al.* (1999) put it, 'it is our suggestion that action is one of the founding principles of our knowledge of the world'. This strongly suggests that they endorse the enactivist account of the visual perception of objects. If so, then it would seem that, for them, what makes processes of embodied simulation *embodied* is precisely that they *are* motor processes.

---

[25] See Milner and Goodale, 1995 and Jacob and Jeannerod, 2003 for slightly different versions of the two-visual systems model of human vision. The fact that some apraxic patients, who cannot use a tool correctly, are not impaired in their conceptual knowledge of tools, is also difficult to reconcile with a motor theory of the conceptual representations of tools.

[26] In Jacob, 2006, I argue that the enactive conception of perception is hard to reconcile with the evidence for the two-visual systems model of human vision.

[27] As I said in section 1, the perception of a graspable object triggers the activity of 'canonical' neurons located in the ventral premotor cortex.

[28] For discussion of Gallagher's enactive account of the perception of another's action, see my criticism in section 4.2 of Blakemore and Decety's (2001) appeal to forward models of action.

On the other hand, Gallese (2003, p. 169; 2004) is very explicit that 'embodied simulation' should not be 'confined to the domain of motor control', which Goldman (2004) interprets as implying that mirroring processes are not restricted to motor cognition. Gallese (2003, 2004) mentions two kinds of evidence for the claim that not all mirroring processes are motor processes, only one of which is relevant here. The relevant piece of evidence is based on brain imaging in humans showing that the same brain area involved in the first-person experience of some basic emotion (e.g. disgust) is also active when an observer recognizes the facial expression of another person experiencing the same emotion.[29] The other piece of evidence—which in my view is irrelevant—is that mental visual imagery does not involve motor processes. This is true, and relevant to the claim that not all mental simulation processes are motor processes. However, it is irrelevant to the claim that not all mirroring processes are motor processes, since mental visual imagery is an instance of pretence, not mirroring.

Interestingly, Gallese (2003, p. 168) links his notion of embodied simulation to so-called 'internal forward models of action' used by computational theorists of action (see section 2.1). On this construal, embodied simulation might be what enables me to predict, in accordance with forward models of action, the sensory consequences of my 'impending act' of reaching for and grasping a glass of water (e.g. that if I lift the glass of water lying on the table, the glass will move upwards and my present visual representation of the glass as lying on the table will need updating). But as I shall argue in the last section, in accordance with forward models of action, embodied simulation so conceived could enable an agent to predict the sensory consequences of his own impending acts only if he is aware of his own motor command by means of a so-called 'efference copy' (of his motor instruction).[30] Unlike an agent, however, an observer is not aware of the agent's motor instruction by means of an efference copy. If so, then it is far from clear how embodied simulation could also be the process whereby the activity of MNs in an observer's brain generates predictions about the sensory consequences of another agent's impending motor act.

2.5 The Distinction Between Lower-Level and Higher-Level Mindreading Tasks

The alternative strategy, recommended by Goldman (2006), is to stick to the pretence-based ST assumption that what underlies mindreading is psychological similarity. On this strategy, the relevant differences between pretence and mirroring can be accommodated by treating pretence as a so-called 'higher-level' mindreading

---

[29] I shall not evaluate the relevance of this evidence for the claim that not all mirroring processes are motor processes here. But so far, I remain unconvinced. For a detailed analysis, cf. Goldman and Sripada, 2005 and Gallese *et al.*, 2004 and Gallese's (2004) discussion at http://www.interdisciplines.org/mirror/papers/1

[30] For details, see section 4.2.

process and mirroring as a 'lower-level' one. Arguably, the distinction between higher-level and lower-level mindreading processes depends on the extent to which the success of a mindreading task requires what Goldman (2006) calls 'quarantining' one's own mental representations or what Leslie (2000), Leslie and Polizzi (1998) and Saxe *et al.* (2004) call 'inhibitory control'. The more inhibitory control is required, the more a mindreading task will count as a higher-level task, and the less inhibitory processing is required, the more it will count as a lower-level task.[31] For example, providing a correct answer to many so-called 'false belief' tasks requires inhibiting one's own correct belief (about e.g. the location of some relevant object) (cf. Bloom and German, 2000 and Hauser, 2003).

To some extent, what is known about the brain areas involved in mindreading and the development of mindreading abilities in children corroborates this conceptual distinction. Evidence from brain imaging in healthy human adults and autistic individuals suggests that reasoning about beliefs and representing goals, intentions and emotions are subserved by different brain areas (cf. Saxe, 2005 and Saxe *et al.*, 2004). Evidence from developmental psychology suggests that toddlers and even infants can represent an agent's immediate goal, which Searle (1983) calls his 'intention in action' and which I shall call his 'motor intention'. Although very recent evidence from Onishi and Baillargeon (2005) suggests that 15-month-old infants can represent an agent's false beliefs, it seems that much younger infants can represent an agent's goal or (motor) intention. For example, Woodward (1998) and Woodward *et al.* (2001) habituated five- to eight-month-old infants to seeing a reach-and-grasp hand movement. The fact that the infants looked longer when the target of prehension changed than when the path of the hand movement changed suggests that the infants' perceptual representation of a human voluntary transitive action involves a selective slot for the target of the action.

According to Saxe *et al.* (2004, pp. 104-105), much of the time when toddlers explicitly represent an agent's volitional state (intention or desire), they rely on their own knowledge of the world and do not represent the agent's beliefs about the world. Although healthy human adults can readily explain a transitive (target-directed) hand action by representing the agent's (true or false) belief together with his desire and/or intention, it is often possible to make sense of the observed action by relying on one's own current perception of the world and representing the goal of the action or the agent's motor intention without explicitly representing the agent's true belief. Whether or not a child endorses the content of an agent's volitional state, the agent's observed movements towards a target, together with the agent's eye-direction, can be taken as reliable cues to the object of his volitional state. Arguably, there is no such cue to the content of an agent's beliefs. If so, then representing an agent's (motor) intention (directed towards a target)—as opposed to representing an agent's false belief—would count as a lower-level mindreading task. Suppose that pretence serves higher-level mindreading tasks and that mirroring

---

[31]   On this account, the distinction is not a classificatory one, but a matter of degree.

serves lower-level tasks. Then the resulting overall picture of simulation-based mindreading processes looks as in Figure 1:
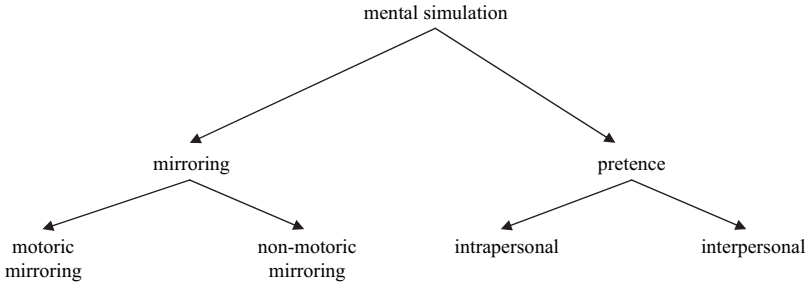
Figure 1 *Goldman's (2006) picture of mental simulation*[32]

Representing an agent's intention might be classified as a lower-level mindreading task in comparison with representing her belief. If the activity of MNs in an observer's brain enables an observer to represent the agent's intention by internally replicating the agent's observed movements, then mirroring would contribute to some third-person mindreading tasks. The question then arises whether motor resonance is sufficient or even necessary for representing an agent's intention.

3. Representing an Agent's Intentions

Since different mindreading tasks make different computational demands, it is conceivable that different tasks are performed by different mindreading capacities. For example, as noticed above, representing an agent's false belief about the location of an object requires inhibiting one's own true belief. As Umiltà's (2001) experiment (discussed in section 1) shows, when a monkey perceives a hand action, MNs in its premotor cortex are only active if a target is present and the monkey knows it. As Saxe (2005) has argued, the fact that MNs remain silent when a target is missing and the monkey knows it shows that the activity of MNs is unsuitable, in these circumstances, for the task of representing the content of an agent's false belief that there is a target. Plausibly however, the ability to represent an agent's false belief should not be taken as a pre-requisite for the ability to represent her intention. Since representing an agent's intention is a different task from representing her false belief (if she has one), MN activity could still enable an observer to represent an agent's intention.

Now, to see what is involved in representing an agent's intention, suppose that I observe a conspecific reach for and grasp a red apple with her right hand: which

---

[32] Goldman (2004, 2006) also assumes that not all mirroring is motor mirroring for reasons having to do with the role of mirroring in recognizing others' emotions. But I cannot examine this further claim here.

constituents of the observed action are likely to be represented by a process of internal replication (or motor simulation)? In grasping the red apple with her right hand, was the agent's goal to eat it? To give it to her little daughter? To throw it away? Or to display it in order to draw it? Suppose that, by matching the agent's hand movement onto my motor system, I can mentally rehearse her reach-to-grasp movement: this might enable me to know what it is like for the agent to grasp and feel the apple within the palm of her right hand. As this example illustrates, the notion of a *goal* is more abstract than the notion of a *target*. Perceiving the target of an agent's act of grasping is a cue—but no more than a cue—to representing the agent's goal. So the question is: could the activity of MNs in an observer's brain generate a representation of the fact that the agent's goal was e.g. to display the apple in order to draw it (as opposed to eating it, giving it to her little daughter, or throwing it away)?

3.1 The Hierarchical Structure of Intentions

Philosophers have long recognized the hierarchical structure of action and drawn a distinction between basic and non-basic acts (see, e.g. Goldman, 1970). Unlike a basic act, a non-basic act is one about which it makes sense to ask *how* the agent performed it. For example, the non-basic action of killing a victim could be performed by several distinct more basic acts, such as pressing the trigger of a gun aimed at the victim, or dropping a poison into the victim's glass of wine.[33] Following Anscombe (1957) and Searle (1983), philosophers have also distinguished the direction of fit of beliefs from that of intentions: beliefs have a mind-to-world direction of fit and intentions a world-to-mind direction of fit. A belief is true if the state of affairs represented obtains (or is a fact) and false otherwise. By contrast, an intention represents a possible non-actual state of affairs: the intention will be satisfied if the agent's action turns the possible state of affairs into a fact.

Typically, an agent does not have one and only one intention in performing an action: rather, she has a network of nested intentions. For example, suppose that an agent has what philosophers, following Searle (1983), call the 'prior' intention to turn the light on, because she wants to read a book and she believes that unless she turns the light on she will not have enough light to read. Turning the light on is a non-basic act since it can be achieved by means of several distinct more basic motor acts (including, for instance, uttering an English sentence requesting an English-speaking conspecific to do it). Let us suppose that, in the circumstances, the agent, who has the prior intention to turn the light on, forms what I shall call the 'motor' intention to press the electrical switch down with her right index finger. By pressing the switch, she will turn the light on. An agent's motor intention

---

[33] Representing the kinematic properties of a bodily movement is representing a lower-level property of an act than representing the goal of an action. See Jeannerod's (1994, 1997) distinction between the semantic and pragmatic representations of an action and see also Jacob and Jeannerod, 2003.

is in essence a motor command (or instruction): it stands to the agent's prior intention in roughly the same relation as her more basic motor act stands to her less basic act.

The agent's prior intention arises from her beliefs and desires: unless it had some conceptual content, it could not interact appropriately with her beliefs and desires. By contrast, the job of an agent's motor intention is to guide and monitor her bodily movements. So an agent's motor intention, unlike her prior intention, has non-conceptual content. Clearly, one and the same motor intention e.g. to press the electrical switch down with one's right index finger, could be at the service of different (even incompatible) prior intentions. For example, depending on the agent's beliefs about the circumstances (such as whether the light is on or off), one and the same motor intention could serve the agent's prior intention either to turn the light *on* or to turn it *off*. In other words, motor intentions stand to prior intentions in a one-to-many relation. Whereas an agent is aware of her prior intention, she is hardly aware of her motor intention.

The need to recognize the hierarchical structure of human intentions is made more pressing by the fact that not all human actions are directed towards inanimate targets: some are directed towards conspecifics. So as well as distinguishing between an agent's motor and prior intentions, we must distinguish an agent's non-social intentions from her social intentions, i.e. her intentions to act towards conspecifics, which, unlike inanimate targets, can act back. What I call a 'social intention' is an intention to affect a conspecific's behavior. Since humans often act out of their mental representations, a social intention may be an intention to modify a conspecific's mental representations. Clearly, much human social cognition (in the sense of Blakemore *et al.*, 2004) depends on the human ability to represent an agent's social intention.

An agent's social intention, like his prior intentions, stand to his motor intentions in a many-one relation. This is illustrated by the case of Dr Jekyll and Mr Hyde (imagined by Jacob and Jeannerod, 2005). Dr Jekyll is a renowned surgeon who performs appendectomies on his anesthetized patients. Mr Hyde is a dangerous sadist who performs exactly the same hand movements on his non-anesthetized victims. It turns out that Mr Hyde is no other than Dr Jekyll. Dr Jekyll alias Mr Hyde may well execute twice the same motor sequence whereby he grasps his scalpel and applies it to the same bodily part of two different persons (one anesthetized, the other suitably paralyzed). If so, then Dr Jekyll's motor intention will match Mr Hyde's. However, Dr Jekyll's social intention clearly differs from Mr Hyde's: whereas the former intends to improve his patient's medical condition, the latter intends to derive pleasure from his victim's pain.

Finally, some of an agent's social intentions are communicative intentions, i.e. as Grice (1989) and others have emphasized, intentions to impart information by virtue of their own recognition by the addressee.[34] For example, consider Jill's

---

[34] See in particular Sperber and Wilson, 1986.

non–verbal communicative intention to inform Bill of her desire to leave the party by ostensively pointing her index finger at her wristwatch in front of him. By recognizing her communicative intention, Bill will acquire the belief that Jill wants to leave the party (and he will only acquire this belief if he does recognize Jill's intention). Jill may, however, execute the very same ostensive bodily movement if she wants Bill to believe instead that her watch is inaccurate. If so, then Jill may hire one and the same motor intention to serve two distinct communicative intentions.

By mentally rehearsing the agent's perceived movement of pressing the switch with her right index finger, an observer could perhaps represent the agent's motor intention. But given that there is no unique mapping from the agent's motor intention to her prior intention, the observer would not thereby be enabled to represent the content of the agent's prior intention. Similarly, by matching onto her own motor repertoire the hand movement whereby Dr Jekyll alias Mr Hyde grasps his scalpel, an observer might be able to represent his motor intention. But this matching would not enable her to discriminate Dr Jekyll's social intention from Mr Hyde's. Finally, by matching Jill's ostensive finger movement onto his own motor repertoire, Bill may represent her motor intention. But again this matching would not enable him to choose which of Jill's two potential communicative intentions her finger movement is supposed to convey.

Furthermore, as argued by Jacob and Jeannerod (2005), evidence from developmental psychology suggests that the matching of an agent's observed movements onto the observer's motor repertoire might not even be *necessary* for representing some of his prior intentions. It has long been known that perceiving the relative motions of geometrical stimuli with no human or animal aspects (e.g. circles and triangles) can prompt normal adults to ascribe emotions and social intentions to the moving stimuli, and describe their motions using intentional verbs such as 'chase', 'corner', 'attack', 'caress', 'comfort' and so on (cf. Heider and Simmel, 1944; Castelli *et al.*, 2000). There is also evidence that 9-month-old infants automatically ascribe goals to moving geometrical stimuli (Gergely *et al.*, 1995; Csibra *et al.*, 1999). Recently, when shown a triangle and a square whose motions were automatically perceived by adults as, respectively, 'helping' and 'hindering' a circle move up a slope, 12-month-old toddlers exhibited a preference for the former over the latter (cf. Kuhlmeier *et al.*, 2003). If so, then it is highly unlikely that human infants represent the intentions of moving geometrical stimuli by a process of motor simulation of (or motor resonance with) the latter's non-biological motion.

Note that while a healthy human agent is fully aware of the conceptual content of her own prior intention, she is hardly aware of the non-conceptual content of her own motor intention (unless a mismatch occurs in the course of her voluntary action). An adult observer may routinely take a representation of the agent's motor intention as input to the computation of hierarchically more abstract representations of the agent's prior, social and/or communicative intentions. Whereas a mindreader is barely aware (if at all) of the content of his representation of the agent's motor

intention, he is acutely aware of the content of his representation of the agent's prior intention. What constitutes the content of the relevant mindreader's belief is the content of his representation of the agent's prior intention, not the content of his representation of the agent's motor intention. If the crucial property of MN activity is resonance, then it can at best deliver a (non-conscious) representation of the agent's motor intention (or motor command), i.e. the tip of the hierarchical iceberg of an agent's nested intentions. Motor resonance on its own lacks the resources to bridge the gap between representing an agent's motor intention and representing her higher-level intentions. The question is: when motor resonance is indeed used, what further cognitive resources will enable a mindreader to move from a non-conscious representation of the agent's motor intention to a conscious representation of her prior intention, and thereby form a belief about the content of that prior intention?

3.2 The New Model of Chains of Logically Related MNs

Recently, MN theorists have addressed some of these questions and designed experiments whose goal is to demonstrate that the activity of MNs extends beyond the mere recognition of a motor act to the representation of an agent's intention. In their own terms, the purpose of the experiments is to show that MN activity enables an observer not merely to respond to the question of *what* the agent is doing, but also of *why* he is doing it by representing what they call (misleadingly, I think) the agent's 'global' intention.[35]

In a series of elegant experiments in which a single motor act is embedded into two more complex actions, Fogassi *et al.* (2005) recorded MNs in the inferior parietal lobule (IPL) of monkeys during the execution and perception of acts of reaching-to-grasp a target in order either to eat it or to place it into a container. In the motor task, the container could be located either near the target or near the animal's mouth. Thus, one and the same act of grasping could be followed by either arm flexion (for eating or placing in a container near the mouth) or arm abduction (for placing near the target). In motor tasks, Fogassi *et al.* (2005) found that the activity of 65% of MNs that fire during execution of grasping is modulated by the more complex action of which it is a part. In particular, MNs fire selectively when grasping is part of placing, whether the placing is near the target or near the mouth (and irrespective of the kinematic differences between arm abduction and arm flexion).

They also recorded MNs in IPL when the animal sees an experimenter grasp a piece of food in order either to eat it or to place it into a container. They found that two thirds of recorded MNs fire selectively during observation of the act of grasping according to whether the grasping was for eating or for

---

[35]   As I argued in the previous section, an agent does not entertain a *global* intention; rather, she entertains a hierarchically organized set of intentions, each of which plays a distinctive causal role in explaining some feature of her behavior.

placing. Finally they found that 84% of MNs exhibit the same specificity in both motor and perceptual tasks. As they recognize, two factors may help the monkey to discriminate between grasping for eating and grasping for placing: one is whether or not the object grasped is food; and the other is whether or not a container is present in the context of the perceived action. However, both are purely perceptual cues, whose processing as such would fail to elicit the activity of MNs.

In an fMRI (imaging study), Iacoboni *et al.* (2005) used the same design and showed healthy human subjects pairs of films divided into three conditions. In the Context condition, subjects saw objects (a teapot, a mug, cookies, etc.) arranged as if either before tea (the 'drinking' Context) or after tea (the 'cleaning' Context). In the Action condition, subjects saw a human hand grasp a mug using either a precision grip or a whole-hand prehension (with no contextual elements present). In the Intention condition, subjects saw one or other of the two acts of prehension embedded in either the drinking context (Intention to drink condition) or the cleaning context (Intention to clean condition).

Compared to a state of rest (i.e. a blank screen), viewing the Intention condition led to the strongest activation in the right inferior frontal areas (known to be rich in MNs). In particular, viewing the Intention condition showed a significant increase of activation in the right frontal areas as compared with viewing the Action condition. Since subjects in the Intention condition (unlike those in the Action condition) perceive many graspable objects other than the grasped mug, this increased activity could arise from the firing of 'canonical' neurons, and not just of MNs. So in order to disentangle the respective contribution of MNs and canonical neurons, the experimenters compared the levels of activity in the Intention to drink condition and the Intention to clean condition. They found that viewing the Intention to drink condition caused a significantly stronger activation of the same brain area than the Intention to clean condition. No such difference was found between viewing the drinking (or before tea) Context and the cleaning (or after tea) Context. Iacoboni *et al.* (2005) conclude that viewing the Intention to drink condition generates a stronger MN activity than viewing the Intention to clean condition.

In the framework of section 3.1, the agent can be said to have the motor intention either to grasp the mug using full-hand prehension or to grasp it using a precision grip. She can also be said to have the prior intention either to drink or to clean. The new results reported by MN theorists raise two fundamental issues. One is whether these experiments show that the activity of MNs generates a representation of an agent's prior intention. The second is whether the observed responses could be generated by a (replicative) process of motor resonance. I will start with the first issue and suggest that it is questionable for two reasons whether these experiments demonstrate that MN activity generates a representation of the agent's prior intention.

Fogassi *et al.*'s (2005) experiment shows that when a monkey perceives a single motor act of grasping, the activity of MNs in the animal's IPL is modulated by his

perception of two factors: whether or not the grasped object is food and whether or not a container is present. Iacoboni *et al.*'s (2005) experiment shows that the presence of contextual cues enhances the activity of inferior frontal areas in humans perceiving an act of grasping a mug. Umiltà *et al.*'s (2001) experiment (reviewed in section 1) showed that when a monkey sees a reach-to-grasp movement, MNs in its premotor cortex fire only if there is a target and the animal knows it. But it did not prove that MN activity is responsible for representing the agent's goal (i.e. to grasp the target), because the representation of this goal could result from an independent purely perceptual representation of the target of the action, based in part on the perceptual representation of the object of the agent's visual attention.

Fogassi *et al.*'s (2005) and Iacoboni *et al.*'s (2005) experiments reveal a correlation between enhanced MN activity and the presence of elements that facilitate a representation of an agent's prior intention. However, correlation is not causation: these experiments do not prove that MN activity *generates* a representation of the agent's prior intention, because this representation might derive from the perceptual processing of contextual cues. Fogassi *et al.* (2005) do not rule out the possibility that perceptual cues (such as whether the object to be grasped is food and whether there is a container) enable the animal to form a perceptual representation of the agent's goal prior to perceiving the act of grasping. Nor do Iacoboni *et al.* (2005) rule out the possibility that the perception of contextual cues gives rise to a perceptual representation of the agent's intention (e.g. to drink), which would be prior to the representation of the agent's motor intention. The enhanced MN activity might itself result from the existence of an independent representation of the agent's prior intention, rather than generating it.

Iacoboni *et al.* (2005) found that the activity in the subjects' right inferior frontal areas showed a significant bias in favor of the Intention to drink condition over the Intention to clean condition. They also report that by making use of their full cognitive capacities (of which the right inferior frontal areas are merely a constituent), subjects could explicitly represent and verbally report the agent's intention to clean as easily as his intention to drink, as a function of contextual cues. Arguably, however, there is a gap between the activity of MNs in the subjects' right inferior frontal areas and their ability to represent a range of alternative intentions as a function of contextual cues. Compared with subjects' full capacity, the very bias of MNs (contained in the inferior frontal areas) for one of the two Intention conditions suggests that their contribution to this capacity is sharply limited.

I now turn to the question of whether strict motor resonance (or contagion) could account for Fogassi *et al.*'s (2005) and Iacoboni *et al.*'s (2005) results.[36] Based on earlier experiments (reviewed in section 1), MN activity was supposed to enable an observer to mentally match the agent's observed movement onto his own motor repertoire. In other words, the most fundamental property of MNs

---

[36] I am grateful to the anonymous referee's request for clarification of my claim that the new model based on chains of logically related MNs is not easy to reconcile with the resonance model of MNs (or the strict congruence between their perceptual and motor properties).

was taken to be the strict congruence between their motor and perceptual properties. The same MNs were supposed to fire during both the execution and the observation of an act of grasping. But as Iacoboni *et al.* (2005, p. 533) note, 'this property cannot account for the present findings, specifically, the difference in responses observed between the drinking and cleaning Intention clips'. Strict congruence (or motor resonance) would require that the same MNs that fire during the execution of a motor act of grasping also fire during the observation of an act of grasping performed by another. However, what the recent experiments show is that MNs that fire in response to the observation of an act of grasping do not fire during the execution of the same motor act: rather, they govern the execution of what Iacoboni *et al.* (2005, p. 533) call a 'functionally related act' (e.g. drinking). Instead of mere replication (strict congruence or motor resonance), MN theorists now appeal to the idea of *chains* of 'logically related' MNs. As Iacoboni *et al.* (2005, p. 533) make clear, however, the relation between pairs of MNs in a chain is probabilistic (or inductive), not strictly logical:

> … the present findings strongly suggest that coding the intention associated with the actions of others is based on the activation of a neuronal chain formed by mirror neurons coding the observed motor act and by 'logically related' mirror neurons coding the motor acts that are most likely to follow the observed one, given the context.[37]

To accept the new model based on chains of logically related MNs is to give up the assumption of strong congruence between the motor and perceptual properties of MNs. In such a chain, MNs coding an observed motor act are supposedly linked to 'logically related' (or probabilistically related) MNs coding the motor act that is most likely to follow its observed immediate predecessor in a given context. In other words, MNs in an observer's brain do not strictly speaking resonate with MNs in an agent's brain: while the latter control the execution of an act of grasping, the former control an act of drinking. Similarly, on the natural assumption that genuine motor contagion requires strict congruence between the perceptual and motor properties of MNs, acceptance of the new model based on chains of logically related MNs generates a discrepancy between motor contagion and the activity of MNs.

Blakemore and Frith (2005) report an interesting experiment in which they requested human subjects to execute sinusoidal movements with their right arm while observing simultaneous arm movements produced by either a human or a robot facing the subjects that were either congruent or incongruent with subjects' own movements. Blakemore and Frith (2005) found that only observation of human (i.e. biological) incongruent arm movements interfered with subjects' own

---

[37]   Cf. Rizzolatti and Craighero, 2005. In order to account for the broad congruence between the perceptual and the motor properties of some MNs, Di Pellegrino *et al.* (1992) already introduced the concept of 'logically related MNs'.

arm movements (by generating variance with respect to a base line condition). They hypothesize that this interference might be evidence for an underlying process of motor contagion. Since the relevant arm movements in this experiment are *not* goal directed, if motor contagion is the source of the interference, it is likely to relate to the kinematic properties of observed and executed bodily movements. The experiment does not show that motor contagion underlies action understanding, i.e. the representation of an agent's goal, let alone of her prior intention. By contrast, activation of a chain of logically related MNs does not seem to constitute a process of genuine motor contagion. Instead of strictly resonating, a chain of logically related MNs constitutes a *predictive* mechanism, whose task is to enable the observer to predict the agent's most likely next act on the combined basis of the observed motor act and contextual cues. Its task is not to enable an observer to mentally replicate the agent's observed motor act.[38]

4. Inverse versus Forward Internal Models of Action

MN theorists have recognized the intrinsic limitations of strict motor resonance as a mechanism for representing an agent's prior intention: motor resonance could at best generate a representation of an agent's motor intention (i.e. motor command). They have thus given up on the requirement of strict congruence between the motor and perceptual properties of MNs. Since they still adhere to Gallese and Goldman's (1998) conjecture that MN activity underlies mindreading (by enabling an observer to represent the agent's intention), they have been led to endorse the new model based on chains of logically related MNs. But on the new model, MN activity is predictive and one cannot endorse it without giving up the strict resonance (or congruence) model of MN activity. I agree that MN activity is predictive, but the question to be addressed in this section is: what do MNs predict?

4.1 Prediction versus Retrodiction

There is a recognizable tension between the new model based on chains of logically related MNs and a critical feature of Gallese and Goldman's (1998) conjecture. According to Gallese and Goldman (1998, p. 497), the 'internally generated

---

[38]   There is further evidence that is hard to reconcile with the view that the main property of neuronal activity in the human ventral premotor cortex is strict motor resonance. For example, Schubotz and von Cramon (2004) found activations in the ventral premotor cortex when subjects saw static arrays of geometrical objects and were asked whether the arrays followed some regular order or not. The subjects' motor system could not resonate with static geometrical stimuli. Furthermore, activations of the human premotor cortex have been reported by Costantini *et al.* (2005), not only for the observation of bio-mechanically possible movements that the subject could perform, but also for bio-mechanically impossible movements. Finally, motor facilitation effects during the observation of impossible movements have been reported by Romani *et al.* (2005).

activation' of MNs in an agent's brain constitutes 'a plan to execute a certain action'. By contrast, the 'external activation' of MNs is supposed to enable an observer to *retrodict* 'the target's mental state' (i.e. her goal or intention) by 'moving backwards from the observed action'. The reason why MN activity could only enable an observer to retrodict a representation of an agent's intention is that an agent's intention is the psychological cause of her movements and hence MN activity could only compute a representation of the agent's intention by working backwards from the perception of her observable movements. But one cannot have it both ways: either MN activity is predictive (in accordance with the new model of logically related MNs) or it is retrodictive (in accordance with Gallese and Goldman's conjecture). In accordance with the new model, it is, I think, more plausible to choose the former option, on which MN activity is predictive.

As noted above, MNs were first discovered in the premotor cortex of monkeys. If they engage in third-person mindreading tasks at all, monkeys are presumably performing lower-level tasks. If and when a non-human primate engages in a one-to-one social interaction with a conspecific, predicting the conspecific's next move should be a relatively automatic process that directly contributes to the agent's decision about which step to take next in her own course of action. Accordingly, when considering the potential contribution of mindreading to the inclusive fitness of non-human primates, Gallese and Goldman (1998, pp. 495-6) emphasize the value of anticipating and predicting a conspecific's next move 'which might be cooperative, non-cooperative, or even threatening'. They do not emphasize the value of retrodicting the conspecific's prior intention. Arguably, retrodicting an agent's prior intention from the perception of his action requires a higher-level capacity to inhibit one's own current action. But it is highly dubious whether non-human primates have such inhibitory capacities.

4.2 Computing an Agent's Intention in Accordance with a Forward Model

In order for an agent to execute a voluntary action, he must have the resources to represent both his prior intention and the motor command suitable to achieving his prior intention. In order to understand the agent's action, an observer must also represent the agent's prior intention and the motor command (or sequence) used by the agent to achieve his goal.[39] Suppose an agent plans and executes a voluntary reaching-to-grasp hand action directed towards a target. If we give up the assumption of strict congruence between the motor and perceptual properties of MNs, the question arises: what brain structures does the agent share with a conspecific observing him?

---

[39] On Dretske's (1988) componential view of behavior, a goal-directed action is the process whereby the agent's intention causes some bodily motor sequence. On this view, a mindreader could either predict the motor sequence from a representation of the agent's intention or retrodict the agent's intention from the agent's observed motor sequence.

What computational theorists of action (Blakemore and Decety, 2001; Blakemore *et al.*, 2000; Miall, 2003; Wolpert, 1997; Wolpert and Ghahramani, 2000; Wolpert *et al.*, 2001) call 'internal models' of action allow us to make some relevant distinctions between the computational resources available, respectively, to an agent and to an observer. According to the internal models theory, during the execution of a voluntary (or intentional) action, an agent's motor cortex simultaneously sends a motor command to the relevant muscles and a copy of this motor command to a central system called a 'comparator' or a 'motor planner'. Internal model theorists of action call the copy of the motor command a 'corollary discharge' or an 'efference copy'. Clearly, the efference copy of the motor instruction is available to the agent, but not to the observer: only the agent is informed of his own agency by the efference copy.[40]

An agent can, but presumably an observer cannot, take the efference copy of his own motor instruction as input and use a *forward* internal model of his own action to compute and predict the sensory consequences of his own motor act. Before executing his reaching-to-grasp movement, an agent can predict, from the efference copy of his own motor command, that if and when he moves his right hand forward, grasps the glass with his right hand and lifts it off the table, his visual experience will represent the glass as being above the table, not on it. In other words, this computation takes a copy of a motor representation as input and computes a perceptual representation.

Since the seminal work of von Holst and Mittelstaedt (1950), forward models of action have been used to explain an agent's ability to distinguish the sensory consequences of his own self-generated movements from sensory changes that are exogenously produced in his environment. They also account for the attenuated experience of the sensory consequences of one's own actions, as compared to the sensory experience of exogenous changes in one's environment: being predictable (from the efference copy of one's motor instructions), the sensory consequences of one's own actions are worth less perceptual attention than sensory changes exogenously produced (which are not predictable).[41] Predicting the sensory consequences of one's own action is representing one's immediate goal. Thus, by using a forward internal model of his own action an agent can compute a sensory (or perceptual) representation of the immediate goal of his action, from the representation of his own motor command (coded by the efference copy).[42]

---

[40]   As already mentioned in section 2.4. Much of the internal models of action framework was anticipated by von Holst and Mittelstaedt (1950), resurrecting ideas of the great vision scientist von Helmholtz, whose concern was to explain how the human brain is able to discriminate the visual signals produced on the retina respectively by the motion of some external moving object and by the endogenously produced movements of the eye.

[41]   This is why one can't tickle oneself, cf. Blakemore *et al.*, 2000.

[42]   Miall (2003) has explicitly considered how MNs might mimic internal forward models of action.

Along similar lines to Gallese and Goldman (1998), Blakemore and Decety (2001) hypothesize that an observer could use a forward internal model of action to retrodict an agent's intention from his observed movements. The observer would first map the agent's observed movements onto her own set of stored predictions of the sensory consequences of her own motor commands. Then, moving backwards, she would estimate what motor commands and intention are likely to have generated the observed movements. Finally, she would attribute these motor commands and this intention to the agent.[43] As noticed by Csibra (2005), not only would this retrodictive account reverse the flow of information typical of the computations performed by forward models, but it also severely under-estimates the computational complexity of mapping an observed movement onto an underlying intention. As I argued in section 3.1, one and the same observable movement can serve many different underlying intentions. Of course, lacking an efference copy, an observer, unlike the agent, cannot directly compute the sensory consequences of the agent's observed action from the efference copy of the agent's motor command. Thus, on this view, both the internally generated activation of MNs in an agent's brain and the externally generated activation of MNs in an observer's brain would contribute to representing the agent's goal; however, the computation would be predictive in the former case, and retrodictive in the latter case.

4.3 Computing Motor Commands in Accordance
with an Inverse Model

As noted above, to have a goal is to represent a non-actual state of affairs, which will be turned into an actual state of affairs by one's action. It might seem that having a goal is a primitive capacity. However, recent work in cognitive neuroscience showing that mental motor imagery can be dissociated from mental visual imagery suggests that having a goal may involve not just one, but two abilities: the ability to represent the motor command required to perform an act (a motor representation) and the ability to represent the sensory consequences of the act (a perceptual representation). For example, Sirigu and Duhamel (2001) instructed healthy subjects, who kept their own hands either on their lap or behind their back, to imagine rotating either their own left hand (first-person condition) or the experimenter's left hand (third-person condition). Mentally rotating one's own hand is a motor imagery task; mentally rotating another's hand is a visual imagery task. When questioned about the spatial position of the little finger on the relevant hand after the imagined rotations, subjects responded significantly faster in the first-person condition when their hands were on their lap (rather than behind their back) and, conversely, in the third-person condition when their hands were

---

[43] This model might make sense of Gallagher's (2007) endorsement of an enactive conception of the perception of human action discussed in section 2.4.

behind their back (rather than on their lap). Thus, in healthy subjects, processes of motor and visual imagery can be selectively facilitated (or impaired).[44]

Not only are motor imagery and visual imagery dissociable, but the theory of internal models of action also enables us to make sense of the twofold nature of goals in terms of the computational contrast between forward and inverse models. Forward models are used to compute the sensory consequences of the agent's action from a representation of the motor command, while inverse models are used to compute motor commands suitable to achieving the agent's prior intention, given, of course, a prior representation of the agent's prior intention.

Now, as MN theorists have emphasized, the fact that MN activity generates motor representations of actions does not in itself make MNs suitable for computing the sensory consequences of actions. Rather, as Csibra (2005) has insightfully argued, it makes them suitable for computing a representation of the motor command required to satisfy the agent's prior intention, in accordance with an inverse internal model of action. If so, then it becomes easier to offer a unified account of the computation performed by MNs despite the basic asymmetry created by the fact that an observer, unlike an agent, lacks an efference copy of the agent's motor instruction. In accordance with inverse models of action, MNs in an agent's brain would compute the motor commands suitable to achieving the agent's prior intention, based on an internally generated representation of this prior intention. In accordance with inverse models of actions, MNs in an observer's brain would also compute a representation of the motor commands that would enable the agent to achieve his prior intention, based on a representation of this prior intention derived from perceptual cues (including a representation of the target of the agent's visual attention).

If the activity of MNs takes a representation of an agent's prior intention as input and computes a representation of motor commands suitable to achieving this goal, then, of course, it does not generate a representation of the agent's prior intention. Rather, the former presupposes the latter. But if so, then, in accordance with Gallese and Goldman's (1998, p. 496) recommendation (quoted above), MNs do enable an observer to predict and anticipate the agent's next move, i.e. his next motor sequence. Arguably, in computing an agent's next move (in accordance with an inverse model of action and consistent with the lack of an efference copy of the agent's motor instructions), the activity of MNs might generate motor imagery (as opposed to visual imagery) in the observer's brain.[45]

On the view of MN activity that I am endorsing, the representation of an agent's prior intention, which serves as input to the inverse model, might arise in an observer's brain as a result of the activity of the STS, which is reciprocally connected

---

[44] Sirigu and Duhamel (2001) further show that patients with, respectively, a parietal lesion and a lesion in the inferotemporal structures exhibit opposite impairments in the hand rotation task, according to whether the task depends on processes of motor or visual imagery.

[45] If so, then this speculation would make sense of Gallese *et al.*'s (2004) otherwise puzzling claim that the activity of MNs underlies *an experiential* grasp of others' minds.

to the inferior parietal cortex, and is known to contain neurons with purely perceptual and no motor properties. This approach to the perceptual contribution of the STS to the representation of an agent's goals and intentions is consistent with much evidence from single cell recordings in monkeys and brain imaging in humans, which shows that neurons in the STS respond to a vast array of eye- and head-movements and movements of the whole body (including locomotion). In light of the critical importance of the perception of the head- and eye-movements of others for tracking their attention and detecting their prior intentions, Allison *et al.* (2000) have suggested that there may be a purely perceptual network of 'social perception' involving the STS, the amygdala and the orbito-frontal cortex.[46] So far, MNs have not been discovered for the execution and perception of head- and eye-movements. If I am right, this is to be expected: the job of MN activity is to compute motor commands from a representation of the agent's prior intention, not to represent the agent's prior intention on the basis of a representation of contextual cues, including the target of the agent's visual attention.

Conclusions

In this paper, I have taken a critical look at Gallese and Goldman's (1998) influential conjecture that MNs constitute a primitive version of a mental simulation heuristic that itself underlies mindreading. Given that MNs were first discovered in macaque monkeys and that the evidence for the existence of mindreading in monkeys is not overwhelming, this is a bold and interesting hypothesis. I have examined what seems to me the most coherent interpretation of this conjecture, namely Goldman's (2006) well-articulated view that not only pretence, but also MN activity, is best thought of as a process of mental simulation (in the similarity sense of 'simulation'). This view in turn is based on two presuppositions. The first is that MN activity is a motor resonance process that creates a neural similarity between an observer and an agent; the second is that this process of motor resonance enables the observer to represent the agent's intention.

After emphasizing the hierarchical structure of a human agent's intentions, I have argued that motor resonance could at best generate a representation of the agent's motor intention (or motor command), not of her (social or non-social) prior intention. MN theorists have designed new experiments to support Gallese and Goldman's conjecture. The results of these experiments have led them to endorse a new model, based on chains of logically related MNs. In considering the new model and the evidence that supports it, I made three complementary points.

First, I argued that the experimental results fail to demonstrate that MN activity generates a representation of the agent's prior intention: they are consistent with an alternative interpretation on which MN activity presupposes

---

[46]   See also Adolphs, 1999 and Brothers, 1990 for an early anticipation of this view.

a representation of the agent's intention, which in turn derives from the perception of contextual cues. Secondly, I have argued that the new model is inconsistent with a strict resonance model of the activity of MNs on which the perceptual properties of MNs are strictly congruent with their motor properties. Thirdly, I have further argued that there is a tension between the predictive role assigned to MNs by the new model and the retrodictive process whereby MNs would compute an agent's intention on Gallese and Goldman's explicit model.

Finally, drawing on an insight from Csibra (2004, 2005) and on the computational resources afforded by the theory of internal models of actions, I have suggested an alternative interpretation of the function of MNs. On this account, MNs do not compute a representation of the agent's intention from a representation of her motor command, in accordance with an internal forward model. Instead, they compute a representation of the agent's motor command from a prior representation of the agent's intention, in accordance with an internal inverse model. Thus, the present account does not detract from the significance of MNs for primates' social cognition, since it emphasizes their contribution to an observer's ability to predict a conspecific's next motor act.

<div align="right">

Institut Jean Nicod
EHESS/ENS-DEC/CNRS

</div>

References

Adolphs, R. 1999: Social cognition and the human brain. *Trends in Cognitive Sciences*, 3, 469–79.

Allison, T., Puce, A. and McCarthy, G. 2000: Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4, 267–78.

Anscombe, G.E. 1957: *Intention*. Oxford: Blackwell.

Baron-Cohen, S. 1995: *Mindblindness, an Essay on Autism and Theory of Mind*, Cambridge, MA: MIT Press.

Blakemore, S-J. and Decety, J. 2001: From the perception of action to the understanding of intention. *Nature Neuroscience*, 2, 561–67.

Blakemore, S-J., Wolpert, D. and Frith, C.D. 2000: Why can't you tickle yourself? *NeuroReport*, 11, R11–R16.

Blakemore, S-J., Winston, J. and Frith, U. 2004: Social cognitive neuroscience: where are we heading? *Trends in Cognitive Sciences*, 8, 216–221.

Blakemore, S-J. and Frith, C. 2005: The role of motor contagion in the prediction of action. *Neuropsychologia*, 43, 260–67.

Bloom, P. 1996: Intention, history, and artifact concepts. *Cognition*, 60, 1–29.

Bloom, P. and German, T. 2000: Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77, B25–B31.

Borg, E. 2007: If MNs are the answer, what was the question? *Journal of Consciousness Studies*, 14, 5–19.

Brothers, L. 1990: The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience*, 1, 27–51.

Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G. and Freund, H.-J.. 2001: Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400–404.

Castelli, F., Happé, F., Frith, U. and Frith, C. 2000: Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, 12, 314–25.

Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaso, A., Romani, G.L. and Aglioti, S.M. 2005: Neural systems underlying observation of humanly impossible movements: an fMRI study. *Cerebral Cortex* (electronic publication ahead of print).

Csibra, G. 2004: Mirror neurons and action observation. Is simulation involved? http://www.interdisciplines.org/mirror/papers/4

Csibra, G. 2005: Mirror neurons or emulator neurons? http://mirrorneurons.free.fr

Csibra, G., Gergely, G., Biro, S., Koos, O. and Brockbank, M. 1999: Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition*, 72, 237–67.

Currie, G. 1995: Visual imagery as the simulation of vision. *Mind & Language*, 10, 25–44.

Currie, G. and Ravenscroft, I. 1997: Mental simulation and motor imagery. *Philosophy of Science*, 64, 161–180.

Currie, G. 1998: Pretence, pretending and metarepresentation. *Mind & Language*, 13, 35–55.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. 1992: Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 91, 176–80.

Dretske, F. 1988: *Explaining Behavior*. Cambridge, MA: MIT Press.

Fadiga, L., Fogassi, L., Pavesi, G. and Rizzolatti, G. 1995: Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608–2611.

Ferrari, P.F., Gallese, V., Rizzolatti, G. and Fogassi, L. 2003: Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17, 1703–14.

Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F. and Rizzolatti, G. 2005: Parietal lobe: from action organisation to intention understanding. *Science*, 308, 662–67.

Gallagher, S. 2007: Logical and phenomenological arguments against simulation theory. In D. Hutto and M. Ratcliffe (eds), *Folk Psychology Re-assessed*. Berlin: Springer Publishers.

Gallese, V. 2003: The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society, London B*, 358, 517–28.

Gallese, V. 2004: Intentional attunement. The mirror neuron system and its role in interpersonal relations. http://www.interdisciplines.org/mirror/papers/1

Gallese, V. 2005: Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences* 4, 23–48.

Gallese, V., Craighero, L., Fadiga, L. and Fogassi, L. 1999: Perception through action. *Psyche*, 5(21) http://psyche.cs.monash.edu.au/v5/psyche-5-21-gallese.html

Gallese, V. and Goldman, A.I. 1998: Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, 2, 493–501.

Gallese, V., Keysers, C. and Rizzolatti, G. 2004: A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.

Gallese, V. and Lakoff, G. 2005: The brain's concepts: the role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, 22, 455–479.

Gallese, V. and Metzinger, T. 2003: Motor ontology: the representational reality of goals, actions and selves. *Philosophical Psychology*, 16, 365–88.

Gergely, G., Nadasdy, Z., Csibra, G. and Biro, S. 1995: Taking the intentional stance at 12 months of age. *Cognition*, 56, 165–93.

Goldman, A.I. 1970: *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice Hall Inc.

Goldman, A.I. 1989: Interpretation psychologized. *Mind & Language*, 4, 161–85.

Goldman, A.I. 1992: In defense of the simulation theory. *Mind & Language*, 7, 104–19.

Goldman, A.I. 2004: Mirror systems, social understanding and social cognition. http://www.interdisciplines.org/mirror/papers/3

Goldman, A.I. 2006: *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Goldman, A.I. and Sripada, C.S. 2005: Simulationist models of face-based emotion recognition. *Cognition*, 94, 193–213.

Gopnik, A. and Meltzoff, A. 1997: *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gopnik, A. and Wellman, H. 1992: Why the child's theory of mind really *is* a theory. *Mind & Language*, 7, 145–71.

Gopnik, A. and Wellman, H. 1994: The theory theory. In L. Hirschfeld and S. Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press.

Gordon, R. 1986: Folk psychology as simulation. *Mind & Language*, 1, 158–71.

Gordon, R. 1992: The simulation theory: objections and misconceptions. *Mind & Language*, 7, 11–34.

Grice, P. 1989: *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hauser, M.D. 2003: Knowing about knowing. Dissociations between perception and action systems over evolution and during development. *Annals of the New York Academy of Sciences*, 1, 1–25.

Heal, J. 1986: Replication and functionalism. In J. Butterfield (ed.), *Language, Mind and Logic*. Cambridge: Cambridge University Press.

Heider, F., and Simmel, M. 1944: An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243–59.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C. and Rizzolatti, G. 2005: Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3, 529–535.

Jacob, P. 2006: Why visual experience is likely to resist being enacted. *Psyche* 12, 1, http://www.psyche.cs.monash.edu.au

Jacob, P. and Jeannerod. 2003: *Ways of Seeing, the Scope and Limits of Visual Cognition*. Oxford: Oxford University Press.

Jacob, P. and Jeannerod, M. 2005: The motor theory of social cognition: a critique. *Trends in Cognitive Sciences*, 9, 21–25.

Jeannerod, M. 1994: The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187–245.

Jeannerod, M. 1997: *The Cognitive Neuroscience of Action*. Oxford: Blackwell.

Jeannerod, M. 2001: Simulation of action. A unifying concept for motor cognition. *NeuroImage*, 14, S103–S109.

Keysers, C., Kohler, E., Umiltà, M.A., Fogassi, L., Rizzolatti, G. and Gallese, V. 2003: Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153, 628–636.

Keysers, C. and Perrett, D.I. 2004: The neural correlates of social perception: a Hebbian network perspective. *Trends in Cognitive Sciences*, 8, 501–507.

Kohler, E., Keysers, C., Umiltà, M.A., Fogassi, L., Gallese, V. and Rizzolatti, G. 2002: Hearing sounds, understanding actions; action representation in mirror neurons. *Science*, 297, 846–48.

Kuhlmeier, V., Wynn, K. and Bloom, P. 2003: Attribution of dispositional states by 12 month-olds. *Psychological Science*, 14, 402–408.

Leslie, A. 2000: 'Theory of Mind' as a mechanism of selective attention. In M. Gazzaniga (ed.), *The New Cognitive Neurosciences*. Cambridge, MA: MIT Press.

Leslie, A.M. and Polizzi, P. 1998: Inhibitory processing in the false belief task: two conjectures. *Development Science*, 1, 247–58.

Mahon, B.Z. and Caramazza, A. 2005: The orchestration of the sensory-motor systems: clues from neuropsychology. *Cognitive Neuropsychology*, 22, 480–94.

Miall, R.C. 2003: Connecting mirror neurons and forward models. *Nature Neuroreport*, 14, 1–3.

Milner, A.D. and Goodale, M.A. 1995: *The Visual Brain in Action*. Oxford: Oxford University Press.

Nichols, S. and Stich, S.P. 2003: *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*. Oxford: Oxford University Press.

Noë, A. 2004: *Action in Perception*. Cambridge, MA: MIT Press.

Onishi, K.H. and Baillargeon, R. 2005: Do 15-month-old infants understand false beliefs? *Science*, 308(April), 255–58.

O'Regan, J.K. and Noë, A. 2001: A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–1031.

Pacherie, E. 2000: The content of intentions. *Mind & Language*, 15, 4, 400–432.

Perrett, D.I., Rolls, E.T. and Caan, W. 1982: Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329–42.

Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K. and Ortega, J.E. 1989: Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, 146, 87–113.

Povinelli, D. and Vonk, J. 2004: We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, 19, 1, 1–28.

Provine, R. 1989: Contagious yawning and infant imitation. *Bulletin of the Psychonomic Society*, 27, 2, 125–126.

Quine, W.V.O. 1960: *Word and Object*, Cambridge, MA: MIT Press.

Ramachandran, V.S. 2000: Mirror neurons and imitation learning as the driving force behind 'the great leap forward' in human evolution. Edge. http://www.edge.org/3rd_culture/ramachandran/ramachandran_pl.html

Rizzolatti, G. and Arbib, M.A. 1998: Language within our grasp. *Trends in Neuroscience*, 21, 188–194.

Rizzolatti, G. and Craighero, L. 2004: The mirror-neuron system. *Annual Review of Neuroscience* 27, 169–92.

Rizzolatti, G. and Craighero, L. 2005: Mirror neurons: a neurological approach to empathy. In J-P. Changeux, A. R. Damasio, W. Singer and Y.. Christen (eds), *Neurobiology of Human Values*. Berlin: Springer.

Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. 1995: Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131–141.

Rizzolatti, G., Fogassi, L. and Gallese, V. 2000: Cortical mechanisms subserving object grasping and action recognition: a new view of the cortical motor functions. In M. Gazzaniga (ed.), *The New Cognitive Neuroscience*. Cambridge, MA: MIT Press, pp. 539–552,.

Rizzolatti, G., Fogassi, L. and Gallese, V. 2001: Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–70.

Romani, M., Cesari, P., Urgesi, C., Facchini, S. and Aglioti, S.M. 2005: Motor facilitation of the human cortical-spino system during observation of bio-mechanically. *NeuroImage*, 26, 755–63.

Saxe, R. 2005: Against simulation: the argument from error. *Trends in Cognitive Sciences*, 9, 4, 174–79.

Saxe, R., Carey, S. and Kanwisher, N. 2004: Understanding other minds: linking developmental psychology and functional neuroimaging. *American Review of Psychology*, 55, 87–124.

Schubotz, R.I. and von Cramon, Y. 2004: Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *The Journal of Neuroscience*, 24, 5467–5474.

Searle, J. 1983: *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University press.

Sirigu, A. and Duhamel, J.R. 2001: Motor and visual imagery as two complementary but neurally dissociable mental processes. *Journal of Cognitive Neuroscience*, 13, 910–19.

Sperber, D. 2004: 'Mirror neurons' or 'concept neurons'? http://www.interdisciplines. org/mirror/1

Sperber, D. and Wilson, D. 1986: *Relevance, Communication and Cognition*. Oxford: Blackwell.

Stich, S. and Nichols, S. 1992: Folk psychology: simulation or tacit theory? *Mind & Language*, 7, 35–71.

Tomasello, M. and Call, J. 1997: *Primate Cognition*. New York: Oxford University Press.

Umiltà, M.A., Kohler, E. Gallese, V. Fogassi, L., Fadiga, L., Keysers, C. and Rizzolatti, G. 2001: 'I know what you are doing': a neurophysiological study. *Neuron*, 32, 91–101.

von Holst, E. and Mittelstaedt, H. 1950: Das Reafferenzprinzip. *Naturwissenschaften*, 37, 464–476.

Wolpert, D.M. 1997: Computational approaches to motor control. *Trends in Cognitive Sciences*, 1, 209–16.

Wolpert, D.M. and Ghahramani, Z. 2000: Computational principles of movement neuroscience. *Nature Neuroscience*, 3, 1212–17.

Wolpert, D.M., Ghahramani, Z. and Flanagan, J.R. 2001: Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5, 487–94.

Woodward, A.L. 1998: Infants selectively encode the goal object of an actor's reach. *Cognition*, 69, 1–34.

Woodward, A.L., Sommerville, J.A. and Guajardo, J.J. 2001: How infants make sense of intentional action. In Malle, B.F. Moses, L.J. and Baldwin, D.A. (eds) *Intentions and Intentionality, Foundations of Social Cognition*. Cambridge, MA: MIT Press.