

After all, it's still replication: A reply to Jacob on simulation and mirror neurons

By *Luca Barlassina*,
Università degli Studi di Milano

Abstract: Mindreading is the ability to attribute mental states to other individuals. According to the simulation theory (ST), mindreading is based on the ability the mind has of replicating others' mental states and processes. Mirror neurons (MNs) are a class of neurons that fire both when an agent performs a goal-directed action and when she observes the same type of action performed by another individual. Since MNs appear to form a replicative mechanism in which a portion of the observer's brain replicates the agent's brain, MNs have been considered evidence in favor of ST. Jacob (2008), however, has maintained that the recent discovery of so-called logically related MNs refutes the hypothesis that MNs form a replicative mechanism. In this paper, I argue that, contrary to what is claimed by Jacob, one can accept the existence of logically related MNs and, at the same time, still maintain that the activity of MNs is replicative. It follows that MNs still support ST.

Keywords: Mirror Neurons; Mindreading; Simulation Theory

Introduction

Mindreading is the ability to attribute mental states to other individuals. To wit, if I believe, or judge, that Mary *desires* to visit London, or that she is *disgusted*, or that she *intends* to grasp a mug, I am performing an act of mindreading. How does mindreading work? That is, what are the cognitive mechanisms underlying attributions of mental states? In answering these questions, the simulation theory of mindreading (ST) starts from the commonsensical idea that we understand what other people think, feel, etc. by putting ourselves in their shoes. In the hands of ST, this intuition is formalized into the idea that a mindreader arrives at attributing a mental state by simulating it in her own cognitive system (Currie and Ravenscroft 2002; Gallese and Goldman 1998; Goldman 2006; Gordon 1986; Heal 2003). ST comes in many varieties and there is no unique notion of simulation that is shared by all its supporters. The majority, however, would agree with Goldman (2006) that the core meaning of 'simulation' in the expression 'simulation theory of mindreading' is *replication*. In other words, the simulation

theorists would be happy with the following statement of their basic point: the ability the mind has to understand other minds is based on its ability to replicate others' mental states and processes. Hereafter, I will use 'simulation' and 'replication' (and their derivatives) as interchangeable terms.

ST has been around from the late 80s (Goldman 1989; Gordon 1986; Heal 1986). However, it got off the ground only some ten years later, following the discovery of mirror neurons (MNs) in the ventral premotor area of macaque monkeys (Gallese *et al.* 1996; Rizzolatti *et al.* 1996), and, subsequently, in the human brain (Rizzolatti *et al.* 2001).¹ It is easy to understand why MNs contributed to the success of ST. According to *the classical model of MNs*, MNs are a class of sensorimotor neurons that fire both when an individual performs some kind of goal-directed action (e.g., grasping, bringing to the mouth), and when it observes the same type of action performed by another individual (Gallese *et al.* 1996; Gallese and Goldman 1998). Two things follow from this. First, MNs form a replicative (or simulational) mechanism in which a portion of the observer's brain matches the observed individual's brain.² Thus, MNs give empirical support to the idea that the mind has the capacity to replicate others' mental states and processes. Second, such a replicative activity takes place when an individual is observing another individual performing an action. Since the observation of another individual's actions is typically accompanied by attributions of mental states to that individual, it is reasonable to hypothesize that mental simulation plays a causal role in mindreading.

Needless to say, the classical model of MNs does not constitute conclusive evidence for ST. If MNs form a replicative mechanism, all we are entitled to assert is that a process of mental simulation is correlated with (at least some acts of) mindreading. But, of course, correlation is not causation. More empirical evidence is needed in order to conclude that a process of mental simulation constitutes the causal basis of (some acts of) mindreading. The best place to look for this further evidence is probably the work of Goldman (see especially Goldman 2006, 2008). However, presenting and assessing such evidence goes beyond the aims of the present paper. Here, rather, I want to discuss a more fundamental issue concerning the relationship

between MNs and ST. This issue has been raised in Jacob (2008). In recent years, following the discovery of the so-called logically related MNs, the classical model of MNs described above has been replaced by *the new model of chains of MNs* (Iacoboni *et al.* 2005). According to Jacob, if one endorses the new model of chains of MNs, one has to give up the idea that MNs form a replicative mechanism. If Jacob is right, ST is in trouble. In fact, if MNs do not form a replicative mechanism, how on earth could they ground the idea that mindreading is based on the capacity the mind has to replicate other minds?

In this paper, I will argue that Jacob is wrong. That is, I will argue that even if one abandons the classical model of MNs in favor of the new model of chains of MNs, one can still maintain that MNs form a replicative mechanism. Accordingly, MNs still support ST. To establish that, I will proceed as follows. In the first section, I will present in some detail the classical model of MNs. In the second section, I will carefully illustrate how the discovery of logically related MNs forced a revision of the classical model of MNs and led to the introduction of the new model of chains of MNs. Finally, in the third section, I will give three arguments to the effect that MNs still form a replicative mechanism in the new model of chains of MNs. A last preliminary observation before beginning: Jacob's critical remarks concerning the alleged relation between ST and MNs exclusively focus on action mirroring, altogether ignoring MNs for emotions and sensations.³ In order to keep the discussion as simple as possible, I will do the same. Accordingly, I will conduct the discussion under the supposition that MNs are confined to the action domain.

1. The classical model of MNs

In this section, I will illustrate the classical model of MNs. I will start by introducing some basic conceptual distinctions concerning the structure of action. I will distinguish between movements, motor acts, and complex actions, a distinction whose cognitive counterpart I will take to be the distinction between kinematic instructions, motor intentions and prior intentions. Let me say a couple of things at the outset about these distinctions.

First, these distinctions are far from satisfactory. However, since Jacob uses a somewhat similar conceptual apparatus (see Jacob 2008, p. 205-212), they will prove useful, in the next section, for understanding Jacob's point. Second, since the notions of kinematic instructions, motor intentions and prior intentions are the counterparts, at the cognitive level, of the notions of movements, motor acts, and complex actions, respectively, I will take the liberty of sometimes using one notion and, other times, using its cognitive counterpart, moving back and forth from the cognitive to the descriptive level. In order not to get lost, keep in mind that, *in general*, what holds for X also holds for its cognitive counterpart Y, and vice versa. This is why I will sometimes draw a conclusion about Y on the basis of X.

What distinguishes motor acts from mere movements is that motor acts are goal-directed movements. Take these three cases. (1) I move my left hand to grasp a mug. (2) I do the same movement in absence of any target. (3) I move my right hand to grasp a mug. In (1) and (3), I performed a motor act. In (2), I did a mere movement. Furthermore, in (1) and (3) I performed the same motor act, i.e., grasping, even though in (1) I performed it by moving my left hand and in (3) by moving my right hand. Hence, on the one hand, two identical movements can *not* count as the same motor act; on the other hand, two different movements can count as the same motor act. The cognitive counterpart of the distinction between movements and motor acts is the distinction between kinematic instructions and motor intentions. While a kinematic instruction suffices to represent a movement, a motor act also requires a motor intention. For instance, when I move my left hand to grasp a mug, my cognitive system is not only specifying a set of muscular movements, but also the motor intention of grasping. Moreover, moving my left hand to grasp a mug and moving my right hand to grasp a mug, although associated with different kinematic instructions, are represented by the same motor intention, namely, the intention to grasp.

The same types of relations that obtain between movements (kinematic instructions) and motor acts (motor intentions) also obtain between motor acts (motor intentions) and complex actions (prior intentions). The same motor act can be at the service of different complex actions. For instance, I

can grasp a mug to drink its contents or I can grasp a mug to clear the table. At the cognitive level, the idea takes the following form: the same motor intention can be nested in different prior intentions. For instance, the motor intention of grasping can be nested either in the prior intention to drink or in the prior intention to clear the table. On the other hand, the same complex action can be realized by different motor acts, and, hence, the same prior intention can contain different motor intentions.

Here is the crucial question. When a subject is observing another individual performing a certain complex action (e.g., grasping a mug with her left hand to drink) what are the observer's MNs coding? Are the observer's MNs coding the agent's kinematic instructions, or the agent's motor intention, or the agent's prior intention? According to the classical model of MNs, MNs code the agent's motor intention. In the remainder of this section, I explain the foundations on which this answer was grounded.

As I said earlier, MNs were first discovered in the ventral premotor area of macaque monkeys; more precisely, in the area F5. Some years before this discovery, single-cell recording experiments showed that the activity of F5 neurons is correlated with specific motor acts and not with particular movements (Rizzolatti *et al.* 1988). For instance, some F5 neurons fire when the monkey grasps a piece of food, regardless of whether this motor act is executed with the right or with the left hand. Moreover, F5 neurons selectively fire during the execution of the same movement, according to whether such a movement is part either of one or of another motor act (for a review, see Rizzolatti and Sinigaglia 2008). On this basis, it was concluded that F5 neurons code motor acts and not mere movements. Using the cognitive vocabulary, F5 neurons code motor intentions.

Some years later, the visual properties of F5 neurons were investigated (Gallese *et al.* 1996; Murata *et al.* 1997; Rizzolatti *et al.* 1996). These studies established the existence of two classes of sensorimotor neurons: first, canonical neurons, which fire both when the monkey performs a particular motor act and when it perceives a graspable object; second, MNs, which fire both when the animal performs a particular motor act and when it observes another individual performing the same motor act. MNs were so called

precisely because their visual properties ‘mirror’ their motor properties. Thus, when an observer is observing another individual performing a certain complex action (e.g., grasping a mug with her left hand to drink), the observer’s brain is coding neither the agent’s movements nor the agent’s complex action, but rather the agent’s motor act (in this case, grasping). At the cognitive level, the observer’s MNs are coding the same motor intention coded by the agent’s MNs.

To sum up, the classical model of MNs is constituted by two claims: first, MNs code motor intentions; second, the activation of MNs in the observer replicates the activation of MNs in the agent. In a nutshell, MNs are the neural basis of a process of motor replication (or motor simulation), in which the observer’s brain replicates (or simulates) the agent’s motor intention. Accordingly, the classical model of MNs gives empirical support to the claim made by simulation theorists that the mind has the capacity to simulate other minds.

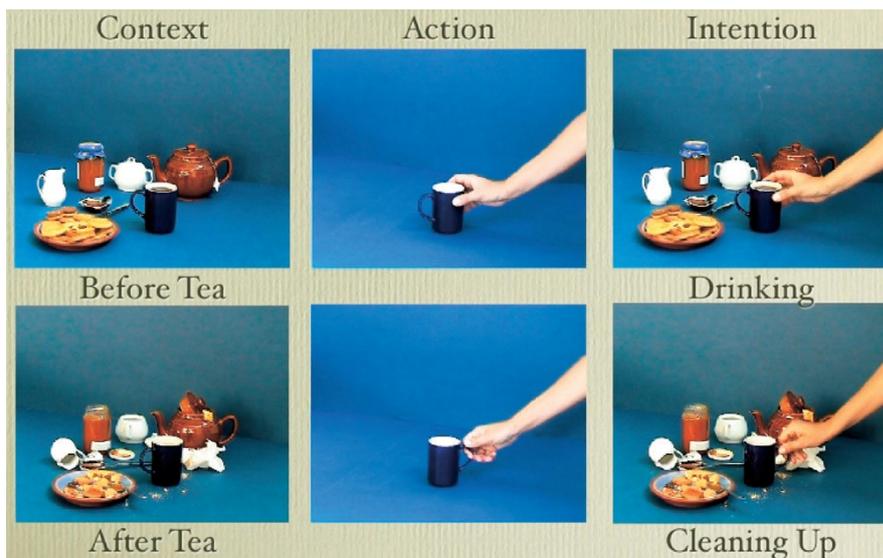
2. *The new model of chains of mirror neurons*

In an important paper published in 2005, Iacoboni and colleagues criticized the classical model of MNs, claiming that MNs not only code motor intentions, but “also play a role in coding the global intention [i.e., the prior intention] of the actor performing a given motor act” (Iacoboni *et al.* 2005, p. 529). Iacoboni *et al.* proposed replacing the classical model of MNs with the new model of chains of MNs, which features the notion of logically related MNs. This proposal was based on the following fMRI study, in which healthy participants observed the same motor act (i.e., grasping) in different contexts.

Subjects were presented with three different types of movie clips, respectively labeled *Context condition*, *Action condition*, and *Intention condition*. The first two types of movie clip (i.e., the Context and the Action conditions) each consisted of two alternative scenes. The Context condition consisted of two scenes with objects related to breakfast. In one of the scenes (the *Before Tea* scene), the objects were arranged as they would be just before breakfast; in the other scene (the *After Tea* scene), as they would be just after. Accordingly, in the Context condition no motor act was depicted. On the other hand, the

second type of movie clip (the Action condition) consisted of two scenes with a hand grasping a mug in absence of any contextual cue, i.e., lacking any other object in the background. In one of the scenes in the Action condition (the *Precision Grip* scene), the hand exerted a precision grip with the fingers on the mug handle; in the other scene (the *Whole-hand Prehension* scene), the hand grasped the body of the mug. Finally, in the third type of movie clip (the Intention condition), both of the scenes represented in the Action condition were embedded in the two scenes that were represented in the Context condition. Accordingly, the Intention condition consisted of 4 alternative scenes. However, for the sake of simplicity, we can ignore the differences between precision grip and whole-hand prehension, and describe the Intention condition as constituted by the motor act ‘grasping’ embedded either in the Before Tea scene or in the After Tea scene. In the former case, we have the *Intention to Drink* scene, since the contextual cues suggest that the motor act of grasping is nested in a complex action guided by the prior intention to drink. In the latter case, we have the *Intention to Clean up* scene, since the motor act of grasping appears to be nested in a complex action guided by the prior intention to clear the table. (The 6 scenes I have just verbally described are depicted in Figure 1).

Fig. 1 (from Iacoboni *et al.* 2005, p. 530)



Now, if, as maintained by the classical model of MNs, MNs only code motor acts (motor intentions), they should fire in the same way when a subject observes the Action and the Intention conditions, since in both conditions one and the same motor act is depicted. However, Iacoboni and colleagues discovered that, compared to the Action condition, the Intention condition yielded significant signal increase in different brain areas, notably in the right inferior frontal cortex, an area that is known to be rich in MNs. Does this show that MNs code prior intentions? Given that one difference between the Intention condition and the Action condition is that in the former, but not in the latter, there are contextual cues that allow deciphering of the prior intention of the agent, one could indeed explain the signal increase in the Intention condition by saying that MNs code prior intentions. However, this conclusion would be premature. In fact, the right inferior frontal cortex also has canonical neurons, i.e., sensorimotor neurons that fire both when an agent performs a particular motor act and when she perceives objects. Thus, one could hypothesize that the greater activation of the right inferior frontal cortex in the Intention condition was not due to MNs, but to canonical neurons activated by the presence of objects that were absent in the Action condition.

In order to adjudicate these two competing hypotheses, Iacoboni and colleagues compared the activation of the right inferior frontal cortex in the Context condition (i.e., the Before Tea and the After Tea scenes) and in the Intention condition (i.e., the Intention to Drink and the Intention to Clean up scenes.) They found that the Before Tea and the After Tea scenes generated a similar activation of the right inferior frontal cortex. However, the observation of the Intention to Drink scene generated a stronger activation in the right inferior frontal cortex than the Intention to Clean up scene. Since the Intention to Drink and the Intention to Clean up scenes depict the same motor act, and since the Before scene and the After Tea scene, in which the Intention to Drink scene and the Intention to Clean up scene are respectively embedded, did not generate differential signal increase, Iacoboni *et al.* ruled out that the differential activation of the right inferior frontal cortex was due either to the motor act or to the contextual objects. Their conclusion was

rather that MNs do not simply code motor acts (e.g., grasping), as maintained by the classical model of MNs, but also code the complex action in which the motor act is nested (e.g., grasping to drink). Using the cognitive vocabulary, the idea can be formulated as follows: MNs do not simply code motor intentions, they also code the prior intention in which the motor intention is nested.

How, according to Iacoboni and colleagues, do MNs code prior intentions? Here is a two-step answer. This answer will introduce us to the new model of chains of MNs. First, Iacoboni and colleagues distinguished two classes of MNs: *classical MNs* and *logically related MNs*.⁴ Classical MNs are simply the good-old fashioned MNs described by the classical model of MNs, that is, neurons firing both when an individual performs a motor act and when she observes the same type of motor act performed by another individual. On the other hand, logically related MNs were not part of the classical model of MNs. They are a new type of MN, described by Iacoboni *et al.* as neurons that do not code the observed motor act but “the motor acts that are most likely to follow the observed one, in a given context” (Iacoboni *et al.* 2005, p. 5). For instance, in the Intention to Drink scene, while the observer’s classical MNs code the motor act ‘grasping’, her logically related MNs code the motor act ‘bringing to the mouth’.

Here is the second part of the story. Iacoboni and colleagues seem to conceive a complex action simply as a chain of more basic motor acts. Analogously, at the cognitive level, a prior intention would reduce to a chain of motor intentions. I have a number of concerns about this way to conceive of the relationship between complex actions (prior intentions) and motor acts (motor intentions), but I will leave such concerns aside. Now, if we assume that a complex action is nothing more than a chain of related motor acts, we can represent the complex actions ‘grasping to drink’ and ‘grasping to clean up’ roughly in the following way:

- (A) complex action of grasping to drink = < (a1) motor act of grasping; (a2) motor act of bringing to the mouth; etc. >
- (B) complex action of grasping to clean up = < (b1) motor act of grasping; (b2) motor act of placing; etc. >

This is the counterpart at the cognitive level:

(A*) prior intention of grasping to drink = < (a1*) motor intention of grasping; (a2*) motor intention of bringing to the mouth; etc. >

(B*) prior intention of grasping to clean up = < (b1*) motor intention of grasping; (b2*) motor intention of placing; etc. >

Putting the two parts of the story together, we get the explanation proposed by Iacoboni and colleagues of how the observer's MNs code the agent's prior intention. As we will see in a bit, such an explanation is based on the notion of chains of MNs. Let's suppose that I am observing the Intention to Drink scene. Since I am observing a motor act of grasping, my classical MNs are coding the motor intention 'grasping'. Given that this motor intention is part of both the prior intentions 'grasping to drink' and 'grasping to clean up' (schematically: a1* is identical with b1*), coding this motor intention is not enough to distinguish these two prior intentions (that is, it is not enough to code A* and not B*). However, the observed motor act is embedded in a context that suggests to me that the next motor act of the agent will be bringing the mug to the mouth. Accordingly, my logically related MNs are coding a2* (i.e., the motor intention 'bringing to the mouth'). Here again, merely coding a2* is not sufficient to code A*. However, if A* is the chain of a1* and a2*, my MNs are globally coding A*. In fact, while my classical MNs are coding a1*, my logically related MNs are coding a2*. Thus, my capacity to code the agent's prior intention "is based on the activation of a *neuronal chain* formed by [classical] MNs coding the observed motor act and by logically related MNs coding the motor acts that are most likely to follow the observed one, in a given context" (Iacoboni *et al.* 2005, p. 5, emphasis mine). This is the central idea of the new model of chains of MNs.

I want to stress a particular characteristic of the new model of chains of MNs, since Jacob's point is based on that characteristic. Let's back to the Intention to Drink scene. The agent is performing a certain motor act (i.e., grasping) and her classical MNs are coding it. At the same time, the observer is seeing the agent performing that motor act. However, the observer's logically related MNs are not coding the agent's observed motor act, but the agent's subsequent motor act (i.e., bringing to the mouth.) According to

Jacob, this discrepancy between what is coded by the agent's classical MNs and what is coded by the observer's logically related MNs is sufficient to conclude that MNs do not form a replicative mechanism. This is how Jacob (2008, p. 211) makes his point:

To accept the new model based on chains of MNs is to give up the assumption of strong congruence between the motor and perceptual properties of MNs. In such a chain, (...) [logically related] MNs in an observer do not strictly speaking resonate with MNs in an agent's brain: while the latter control the execution of grasping, the former control an act of drinking.

As I said earlier, if Jacob is right in saying that MNs do not form a replicative mechanism, then MNs cease to be evidence in favor of ST. In that case, simulation theorists would be in trouble, since MNs are probably the strongest evidence supporting ST. In the next section, however, I will argue that even if one endorses the new model of chains of MNs, one does not need to abandon the idea that MNs form a replicative mechanism.

3. Three replies to Jacob

In what follows, I will give three arguments to the effect that it is possible to accept the new model of chains of MNs and, at the same time, still maintain that MNs form a replicative (or simulational) mechanism. I take these arguments to have different strengths. The first one, I should admit, is only a partial response to Jacob. The second argument, on the other hand, if successful, would count as a full-blown rejection of the case made by Jacob. However, I don't expect Jacob to share the intuition on which it is based. Thus, the crucial argument turns to be the third one, which I think convincingly refutes Jacob's point.

3.1 Partial replication

This is my first argument. Iacoboni *et al.*'s criticism of the classical model of MNs does not amount to denying the existence of classical MNs. Iacoboni and colleagues, rather, claim that in addition to classical MNs, we should introduce logically related MNs. Thus, in the new model of chains of MNs, we have two types of MNs: classical MNs and logically related MNs. Now, classical MNs are conceived in the same way both in the classical model of

MNs and in the new model proposed by Iacoboni and colleagues. In fact, Iacoboni *et al.* still define classical MNs as neurons showing “a congruence between their visual and motor properties” (Iacoboni et al. 2005, p. 533). So, one might reply to Jacob that even if one endorses the new model of chains of MNs, MNs still form a mechanism that is *partially* replicative. More precisely, in the new model of chains of MNs, the subset of MNs constituted by classical MNs forms a replicative mechanism. To understand this point, consider the following case.

An observer, O, is observing an agent, A, in the Intention to Drink scene. That is, O is observing A grasping a mug in a context that suggests that A intends to drink from the cup. If we label ‘time t' ’ the time during which A is grasping the mug, we can say that, at time t , A’s classical MNs are coding the motor act of grasping. True enough, at time t , O’s logically related MNs are coding a different motor act, that is, ‘bringing to the mouth’. Nonetheless, at time t , O’s classical MNs are coding the motor act of grasping. So, at time t , O’s classical MNs are replicating the activity of A’s classical MNs.

To sum up, my first reply to Jacob is that even in the new model of chains of MNs, the activity of MNs is *partially* replicative, since classical MNs still form a replicative mechanism. This reply is not entirely satisfactory, however. Jacob’s aim was differentiating MNs and simulation. By saying that MNs form a mechanism that is *partially* replicative, one is simultaneously conceding that such a mechanism is *partially non-replicative*. So, one is admitting that Jacob was somewhat successful in separating MNs from simulation. In order to give a better reply to Jacob, one should not content oneself with the idea that, in the new model of chains of MNs, the activity of classical MNs is still replicative. One rather should show that, in the new model of chains of MNs, logically related MNs form a replicative mechanism too. This leads us to my second argument.

3.2 Non-synchronic simulation

Let’s look back at the case in which observer O is observing agent A in the Intention to Drink scene. That is, let’s imagine once again that O is observing A grasping a mug in a context that suggests that A intends to drink from the

cup. This time, let's also suppose that A's prior intention really is to drink from the cup. Accordingly, A's next motor act will be bringing the mug to the mouth. If we label as 'time t ' the time during which A is grasping the mug, we can say that at time t O's logically related MNs are not coding A's current motor act, but rather the motor act that A will perform at the time $t + 1$, i.e., bringing to the mouth. Hence Jacob would conclude that O's logically related MNs are not simulating A's brain state. This conclusion deserves a careful scrutiny.

Let's first notice that there is a certain type of *isomorphism* between the brain states of A and O. In fact, at time t , O's logically related MNs are coding a motor intention (i.e., bringing to the mouth) that A's brain is also going to code, albeit not until time $t + 1$. Accordingly, we here have a case of *non-synchronic* isomorphism. So, the following question arises: in order for a process S to simulate a process P, do S and P have to occur synchronically? The right answer seems to be: "no, they don't". For example, consider the following case: one day, Mike was piloting a plane. That day, something went wrong, and Mike was forced to perform an emergency landing. The next day, in order to discover what went wrong, a group of engineers run a flight simulation. I take to be uncontroversial that the flight simulation is a simulation *of* Mike's flight. However, the two events did not occur synchronically. In fact, while Mike's flight took place at day n , the simulation took place at the day $n + 1$. Hence, simulation does not require synchronicity and, accordingly, there is no reason to deny that O's logically related MNs are simulating A's brain state.

Jacob could reply as follows. In the case of the flight, the flight simulation took place after the real flight. In this sense, it was a *replication* of a previous event. But in the case of MNs, matters are the other way round. That is, the observer's brain is one step forward the agent's brain. In fact, at time t O's logically related MNs are coding the motor act that A will perform at time $t + 1$. So, the problem is not whether a simulation S has to occur at the same time of the process P, but rather whether S can precede P. To evaluate this possibility, let's refer back to the flight case. One day, Mike was using a flight simulator to simulate an emergency landing. Let's call this event 'E'. The next

day, Mike flew a real plane. That day, something went wrong, and Mike was forced to perform an emergency landing. Let's call this event 'E*'. My intuition is that even though E temporally preceded E*, it is E that constitutes a simulation of E*, and not vice versa. If my intuition is correct, the fact that at time t O's logically related MNs are coding the motor act that A will perform at time $t + 1$ is not sufficient to conclude that at time t O's logically related MNs are not simulating A's brain state. Rather, logically related MNs instantiate a process of *non-synchronic simulation*.

I should confess, however, that I don't expect my intuition to settle the controversy. In fact, I imagine that Jacob would reply that he has the opposite intuition, i.e., that in order for the relation of simulation to obtain, the two *relata* have to be in a particular temporal relation. In this case, we would reach a deadlock. In order to break up this potential scenario, I am going to propose my third and last argument. As in my second argument, the upshot of this final argument is to establish that logically related MNs also form a replicative mechanism. However, unlike my second argument, my final argument is empirically grounded and does not rely on controversial intuitions. Furthermore, I will claim that logically related MNs instantiate a process of *synchronic simulation*, just as classical MNs do.

3.3 Visual and motor properties of logically related MNs

My third argument is based on a study by Fogassi *et al.* (2005), in which the functional properties of neurons in the inferior parietal lobule (IPL) of monkeys were investigated. The study involved two monkeys and consisted of two parts.

In the first part, the activity of 165 single neurons was studied when the monkey performed the same motor act (i.e., grasping) embedded in different complex actions. In the first condition, the monkey grasped a piece of food in front of itself and brought the food to its mouth (*Grasping to eat* condition). In the second condition, after grasping the food, the monkey placed the food in a container in front of itself (*Grasping to place* condition). So, the motor act of grasping was followed either by the motor act of bringing to the mouth or by the motor act of placing. Fogassi and colleagues found that some IPL neurons

fired in the same way in both the Grasping to eat condition and the Grasping to place condition. However, the large majority of IPL neurons selectively fired *during* the motor act of grasping, according to whether this motor act was followed either by the motor act of bringing to the mouth or by the motor act of placing (that is to say, while some neurons fired *during* the motor act of grasping only when it was followed by the motor act of placing, other neurons fired *during* the motor act of grasping only when it was followed by the motor act of bringing to the mouth).⁵

The first part of Fogassi *et al.*'s experiment established the existence of two types of IPL neurons: the first type fires when an agent is performing a certain motor act, regardless of the subsequent motor act; the second type selectively fires during the performance of a certain motor act, in accordance with the subsequent motor act. More precisely, when an agent *A* is performing a motor act *E* at time *t*, a subset α of *A*'s IPL neurons is coding *E*, while another subset β of *A*'s IPL neurons is coding the next motor act – that is, motor act *E* + 1 – which *A* will perform at time *t* + 1.

Some of the 165 IPL neurons studied by Fogassi and colleagues in the first part of their experiment were already known to be MNs, given that it was earlier established that they fire both when a monkey perform a motor act and when it observes the same motor act performed by another agent (Fogassi *et al.* 1998). In the second part of the experiment, Fogassi *et al.* investigated the visual properties of 41 of these IPL MNs. In order to do so, they set a visual task in which the monkey was presented with an experimenter performing the same two complex actions that the monkey performed in the Grasping to Eat and in the Grasping to Place conditions. The results were as follows: while some MNs fired in the same way during the observation of both of the conditions, the large majority of neurons selectively fired *during* the observation of the motor act of grasping, according to whether this motor act was followed either by the motor act of bringing to the mouth or by the motor act of placing.⁶

Needless to say, what Fogassi *et al.* discovered in the second part of their experiment is in line with the findings of Iacoboni *et al.* (2005). Fogassi and colleagues established the existence of two types of MNs. The first type

discharges when an individual is observing an agent performing a certain motor act, regardless of the agent's subsequent motor act. Hence, this first type of MN is a classical MN, coding the observed agent's current motor act. On the other hand, the second type of MN does not code the observed agent's current motor act, but rather the agent's subsequent motor act. So, these are logically related MNs. Schematically, when an observer O is observing an agent A performing a motor act E at time t , O's classical MNs are coding E, while O's logically related MNs are coding A's next motor act – that is, motor act E + 1 – which A will perform at time $t + 1$.

Recall that the 41 IPL MNs studied in the second part of Fogassi *et al.*'s experiment were members of the set of 165 IPL neurons studied in the first part of the experiment. So while Iacoboni et al. (2005) only investigated the visual properties of logically related MNs, Fogassi *et al.* also investigated their motor properties. By comparing the visual behavior with the motor behavior of these logically related MNs, I will show that Jacob is wrong in claiming that logically related MNs do not form a replicative mechanism.

Of the 41 IPL MNs, 19 were logically related MNs. In the action task, 15 of them fired during the motor act of grasping when it was followed by the motor act of bringing to the mouth (call them 'Grasping to Eat logically related MNs'.) On the other hand, 4 of them fired during the motor act of grasping when it was followed by the motor act of placing (call them 'Grasping to Place logically related MNs'.) So, when an agent A is performing a motor act E at time t , if A's next motor act is bringing to the mouth, A's 'Grasping to Eat logically related MNs' are firing at time t . On the other hand, if A's next motor act is placing, A's 'Grasping to Place logically related MNs' are discharging at time t . What about the visual responsive properties of IPL logically related MNs? Strikingly, 16 out of 19 logically related MNs showed the same selectivity in the observation task. More precisely, 13 out of 15 'Grasping to Eat logically related MNs' also fired when observing the Grasping to Eat condition, and 3 out of 4 'Grasping to Place logically related MNs' also fired when observing the Grasping to Place condition. Thus, 84% of IPL logically related MNs exhibit the same selective behavior in both action and perception.

This strong congruence between the motor and the perceptual properties of logically related MNs clearly speaks against Jacob's idea that logically related MNs do not form a replicative mechanism. When, at time t , an agent A is performing a motor act E, A's logically related MNs are already coding A's next motor act. Analogously, when, at the same time, an observer O is observing A performing E, O's logically related MNs are already coding A's next motor act. In a nutshell, logically related MNs in an observer replicate the activity of logically related MNs in the agent, since they both code the agent's next motor act.

Conclusion

According to ST, the ability the mind has for understanding other minds is based on the ability it has for replicating others' mental states and processes. Classical MNs give empirical support to this idea. In fact, when, at time t , an agent A is performing a motor act E and an observer O is observing E, both A's and O's classical MNs are coding E. However, if one accepts the new model of chains of MNs, one should admit the existence of another type of MNs, namely, logically related MNs. When, at time t , an agent A is performing a motor act E and an observer O is observing E, whereas A's classical MNs are coding E, O's logically related MNs are coding A's next motor act. On this basis, Jacob concluded that accepting the new model of chains of MNs entails giving up the idea that MNs form a replicative mechanism. In this paper, I have shown that Jacob's conclusion is a *non sequitur*. It is true that logically related MNs in an observer don't strictly speaking resonate with classical MNs in the agent's brain. However – and this was my main point – Fogassi et al.'s (2005) study established that the activity of logically related MNs in an observer replicates the activity of logically related MNs in the agent. When, at time t , an agent A is performing a motor act E and an observer O is observing E, both A's and O's logically related MNs are coding A's next motor act. Hence, *pace* Jacob, it is possible to accept the new model of chains of MNs and, at the same time, still maintain that MNs form a replicative mechanism. Thus, MNs still provide empirical support for ST.

Acknowledgements

I am very grateful to Pierre Jacob for the numerous discussions we had on MNs and ST. Even though I disagree with most of Jacob's idea about these topics, I enormously benefited from our exchanges. I am indebted to Elizabeth Oldfather, who read the paper and provided a number of insightful comments. I gave an earlier version of this paper at the workshop on 'Neuroethics and Consciousness', held at the University of Turku, Finland, on August 2010. Thanks to Juha Räikkä for inviting me to the workshop and to all the participants for their very useful comments. I want to express my gratitude to Rosa Rantanen, who nicely guided me through the streets of Turku. Finally, a very special thanks to Susanne Uusitalo, for her kind hospitality and invaluable friendship.

References

- Avenanti, A., Buetti, D., Galati, G. and Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience*, 8: 955-960.
- Currie, G. and Ravenscroft, I. (2002). *Recreative Minds*. Oxford: Oxford University Press.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V. and Rizzolatti, G. (1992). Understanding motor events: a neuropsychological study. *Experimental Brain Research*, 91: 176-180.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. and Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, 308: 662-667.
- Fogassi, L., Gallese, V., Fadiga, L. and Rizzolatti, G. (1998). Neurons responding to the sight of goal-directed hand/arm actions in the parietal area PF of the macaque monkey. *Society for Neuroscience Abstracts*, 24, 257.5.
- Gallese, V., Fadiga, L., Fogassi, L. and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119: 593-609.
- Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mindreading. *Trends in Cognitive Sciences*, 2: 493-501.
- Goldman, A. (2006). *Simulating Minds*. Oxford: Oxford University Press.
- Goldman, A. (2008). Mirroring, mindreading, and simulation. In J. Pineda (ed.), *Mirror Neuron Systems: The Role of Mirroring Processes In Social Cognition*. Humana Press.
- Gordon, R. (1986). Folk psychology as simulation. *Mind & Language*, 1: 158-171.
- Heal, J. (1986). Replication and functionalism. In J. Butterfield (ed.), *Language, Mind, and Logic*. Cambridge: Cambridge University Press.
- Heal, J. (2003). *Mind, Reason and Imagination*. Cambridge: Cambridge University Press.

- 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. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language*, 23: 190-223.
- Keysers, C., Wicker, B., Gazzola, V., Anton, J-L., Fogassi, L. and Gallese, V. (2004). A touching sight: SII/PV activation during the observation of touch. *Neuron*, 42: 335-346.
- Murata, A., Fadiga, L., Fogassi, L. Gallese, V., Raos, V. and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*, 78: 2226-2230.
- Rizzolatti, G., Camarda, F., Fogassi, L., Gentilucci, M., Luppino, G. and Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research*, 71: 491-507.
- Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3: 131-141.
- Rizzolatti, G., Fogassi, L. and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2: 661-670.
- Rizzolatti, G. and Sinigaglia, C. (2008). *Mirrors in the Brain*. Oxford: Oxford University Press.
- Wicker, B., Keysers, C., Plailly, J., Royet, J-P., Gallese, V. and Rizzolatti, G. (2003). Both of us disgusted in *my* insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40: 655-664.

Notes

- ¹The success of ST should probably be traced back to the publication of Gallese and Goldman's (1998) groundbreaking paper.
- ² Along with 'simulation' and 'replication', the expression 'matching' is also widely used in the literature on MNs. Some authors use the expressions 'resonance' or 'rehearsing' as well. For my present aims, we can treat all these expressions as synonymous. For an attempt to analyze some of these expressions and disentangle their meanings, see Goldman (2008).
- ³ See Avenanti *et al.* (2005) for MNs for pain; Keysers *et al.* (2004) for MNs for touch; Wicker *et al.* (2003) for MNs for disgust. For a review, see Goldman (2008).
- ⁴ The expression 'logically related MNs' was originally introduced in Di Pellegrino *et al.* (1992).
- ⁵ To control for the possibility that the differential discharge was not due to the nature of the subsequent motor act, but to the nature of the subsequent movements, Fogassi *et al.* introduced a third condition, in which the same initial motor act was followed by placing the food in a container located on the monkey's shoulder. Thus, in the third condition, the subsequent motor act was identical to the one occurring in the Grasping to Place condition, but in terms of its kinematic properties, it was more similar to the subsequent motor act occurring in the Grasping to Eat condition, since it required an arm flexion, as in the case when the monkey brings the food to the mouth. Fogassi and colleagues discovered that all the neurons that selectively discharged in the Grasping to Place condition discharged in the third condition too. Conversely, the neurons that

discharged in the Grasping to Eat condition did not discharge in the third condition, despite the kinematic similarities.

⁶ Controversy remains as to how the second type of neuron was able to code the subsequent act. I agree with Jacob that “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” (Jacob 2008, p. 209). Jacob draws important theoretical conclusions from the role played by contextual cues in deciphering the subsequent motor act (Jacob 2008, pp. 208-210). In this paper, however, I will not discuss his conclusions.