

What do mirror neurons mirror?

Sebo Uithol, Iris van Rooij, Harold Bekkering and
Pim Haselager

Single cell recordings in monkeys provide strong evidence for an important role of the motor system in action understanding. This evidence is backed up by data from studies of the (human) mirror neuron system using neuroimaging or TMS techniques, and behavioral experiments. Although the data acquired from single cell recordings are generally considered to be robust, several debates have shown that the interpretation of these data is far from straightforward. We will show that research based on single-cell recordings allows for unlimited content attribution to mirror neurons. We will argue that a theoretical analysis of the mirroring process, combined with behavioral and brain studies, can provide the necessary limitations. A complexity analysis of the type of processing attributed to the mirror neuron system can help formulate restrictions on what mirroring is and what cognitive functions could, in principle, be explained by a mirror mechanism. We argue that processing at higher levels of abstraction needs assistance of non-mirroring processes to such an extent that subsuming the processes needed to infer goals from actions under the label ‘mirroring’ is not warranted.

Keywords: Computation; Goal Inference; Mirror Neurons; Representation; Social Cognition; Theory of Mind

1. Mirroring in Cognitive Neuroscience

After their discovery in the early 1990s (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) mirror neurons caused great excitement in cognitive neuroscience, as these neurons seem to suggest a common coding of action perception and action execution (Hommel, Müsseler, Aschersleben, & Prinz, 2001;

Sebo Uithol is a PhD student at the Radboud University Nijmegen, Donders Institute.

Iris van Rooij is Assistant Professor at the Radboud University Nijmegen, Donders Institute.

Harold Bekkering is Full Professor at the Radboud University Nijmegen, Donders Institute.

Pim Haselager is Associate Professor at the Radboud University Nijmegen, Donders Institute.

Correspondence to: Sebo Uithol, Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, PO Box 9104, 6500 HE Nijmegen, The Netherlands. Tel: +31 24 361 6068. Fax: +31 24 361 6066. Email: S.Uithol@donders.ru.nl

Prinz, 1997), or a shared representation between the observer and the executor of an action (de Vignemont & Haggard, 2008; Grèzes & Decety, 2001). They were labeled ‘mirror neurons’ as the observed action “seems to be ‘reflected’, like in a mirror, in the motor representation for the same action of the observer” (Buccino, Binkofski, & Riggio, 2004a, p. 371). It is exactly this “rock bottom” connotation of mirroring (i.e., direct reflection, or direct matching of action features) that has made it an attractive notion for explanations of cognitive functions, including action understanding (Evangelidou, Raos, Galletti, & Savaki, 2009; Filimon, Nelson, Hagler, & Sereno, 2007; Rizzolatti, Fogassi, & Gallese, 2001) emotion understanding (de Vignemont & Singer, 2006; Gallese, 2001; Keysers & Gazzola, 2006; Wicker et al., 2003), imitation (Brass & Heyes, 2005; Buccino et al., 2004b; Iacoboni et al., 1999; Rizzolatti, 2005; Wohlschläger & Bekkering, 2002), complementary action (Newman-Norlund, Van Schie, Van Zuijlen, & Bekkering, 2007), communication (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003; Gallese & Lakoff, 2005), inference of intentions and goals (Fogassi et al., 2005; Iacoboni et al., 2005; Rizzolatti et al., 2001), and social cognition in general (Gallese & Goldman, 1998; Goldman, 2006; Iacoboni & Dapretto, 2006).

Evidence for the existence of mirroring processes is derived from broadly three types of experimental research: (i) single-cell recordings in monkeys, (ii) analyses of the entire (human) mirror neuron system (MNS) using imaging or TMS techniques, and (iii) behavioral experiments, using interference effects and reaction times to probe properties of the MNS. The received view is that observed actions are mapped onto the motor cortex of the observer. When there is a matching motor representation available, the action is recognized. This hypothesis is known as the *direct-matching hypothesis* (Rizzolatti et al., 2001; Rizzolatti & Sinigaglia, 2010). In single cell research, the activity of mirror neurons is often conceptualized as a form of representation, coding for (categories of) actions or action goals (Fogassi et al., 2005; Gallese et al., 1996; Iacoboni et al., 1999; Rizzolatti et al., 1996). In this type of research, the activity of a single neuron is measured and related to the occurrence of an external event. When there is a reliable covariance between the neuronal activity and an external event it is concluded that the neuronal activity represents the external event.

In research based on imaging techniques or TMS and behavioral studies, mirroring is generally viewed as a form of processing, mapping perceptual representations of the observed action to motor representations of the observer’s own action repertoire (Buccino et al., 2004a; Iacoboni et al., 2005; Nakahara & Miyashita, 2005; Rizzolatti, 2005; Rizzolatti & Craighero, 2004; Rizzolatti et al., 2001). As this type of research is dependent on imaging or TMS techniques, reaction times and error rates, which can only show or influence activity in large groups of neurons, one can show the involvement of a brain region as a whole in a certain task, but not the response or contribution of a single neuron.¹

The data acquired from single cell recordings is generally regarded to be robust and solid. Therefore, it is often used to guide research of the other two types, or to interpret the acquired data. For instance, Newman-Norlund et al. (2007) use the distribution of strictly and broadly congruent mirror neurons, as found by Gallese and

colleagues (1996) in monkeys, to predict BOLD signals in the MNS in two different conditions (imitative vs. complementary action). However, the existence of several debates about the function of mirror neurons (Csibra, 2007; Dinstein, Thomas, Behrmann, & Heeger, 2008; Jacob, 2008; Jacob & Jeannerod, 2005; Saxe, 2005) indicate that although the data these single-cell experiments generate might be hard, the interpretation of these findings is far from straightforward.

In recent years several researchers have formulated criticisms on the received interpretation of the function of mirror neurons (the direct-matching hypothesis) pointing to the fact that this hypothesis cannot account for many important findings, and have formulated alternative theories (e.g., Csibra, 2007; de Vignemont & Haggard, 2008; Jacob, 2008). For example, both Csibra and Jacob argue that mirror neuron activity is not constitutive of action understanding, but only indicative of it (Csibra, 2007; Jacob, 2008; Jacob & Jeannerod, 2005). They argue that action understanding is an interpretative process that takes place outside the motor system and that mirror neurons are involved in the subsequent action prediction and planning. Although our analysis is different, the outcome can be interpreted as (partially) supporting their views. In this paper we want to analyze the paradigm that has led to many of these findings, i.e., the single cell recordings. By means of an analysis of the representational elements of mirror neurons, we will show that this type of research allows for virtually unbounded content attribution to individual neurons (see also Uithol, Haselager, & Bekkering, 2008). As a consequence, mirror neurons can be said to represent ever more abstract events, from grip types to long-term intentions. However, by means of a complexity analysis of goal inference, a task generally attributed to a mirroring process, we can formulate a possible limitation on what representational output such a process can produce. This analysis, combined with behavioral or brain studies, provides a possible means of limiting the attributable representational content and can thereby help in interpreting the data acquired with single-cell measurements. We will argue that the recognition and understanding of goals and intentions need assistance of non-mirroring processes to such an extent that subsuming these processes under the label 'mirroring' is no longer warranted. Our analysis is specifically aimed at a mirror mechanism and its alleged support for action understanding. Although there might be consequences for its support for other cognitive functions (see above), these fall outside the scope of this paper.

2. Mirror Neurons Representing Actions and Goals

The firing of mirror neurons is often characterized as a form of representation. The neurons are said to represent action means, action ends or goals, and intentions. Examples are abundant: Gallese et al. propose that a "possible function of mirror neuron movement representation is that this representation is involved in the 'understanding' of motor events" (1996, p. 606); Rizzolatti et al. propose that "[mirror neuron's] activity 'represents' the observed action" (1996, p. 131); and

Iacoboni et al. suggest that “F5 neurons code the general goal of a movement” (1999, p. 2527).

What counts as an action means or an action goal is relative and a matter of interpretation. To give an example: a precision grip can be a means to the grasping of a cup. This cup grasping, however, can also be considered a means to the end of drinking. Drinking, in its place, can be regarded a means to maintaining homeostasis or engage in social activity. There thus exists a continuum from concrete, readily observable movements (e.g., the use of a precision grip) to highly abstract goals and intentions (such as engaging in social activity), and there is no a priori way to make a clear-cut and objective contrast between action means and action ends or goals. Several action hierarchies and labels have been proposed to divide this continuum (Bekkering & Wohlschläger, 2002; Grafton & Hamilton, 2007; Jeannerod, 1994; Pacherie, 2008). To clarify our terminology: we will speak of actions in relation to the level of grips or simple actions (e.g., grasping with a precision grip), and of goals when the behavior is interpreted more broadly, ranging from motor goals (e.g., the goal of grasping a cup) to long-term intentions (cleaning the table or spending your next holiday in Brazil).²

As mirror neuron firing is commonly viewed as a form of representation, we will analyze it using the basic elements of a representation: *vehicle*, *content*, *object* and *user* (Cummins, 1989; Dretske, 1988; see Figure 1; see also Bechtel, 1998; and Shea, 2007, for similar presentations of these elements). The representation proper consists of a vehicle and a content. The *vehicle* of a representation is the physical carrier (e.g., neural state) that represents. The information that is carried by the vehicle is called its *content*. Content is not the same as the *object* that is represented. An object or event in the outside world can be misrepresented and most of the time the content is of a more general or more abstract nature than the object represented (e.g., a sparrow can get represented as “bird”). It is important to note that representational objects need not be physical objects. A representational object can as well be a situation or an event, such as an action. The fourth and final element of a representation is a *user*. The user is the system or process that uses the representation to guide its behavior. In case of mirror neurons, the user is likely to be another brain system. For a full

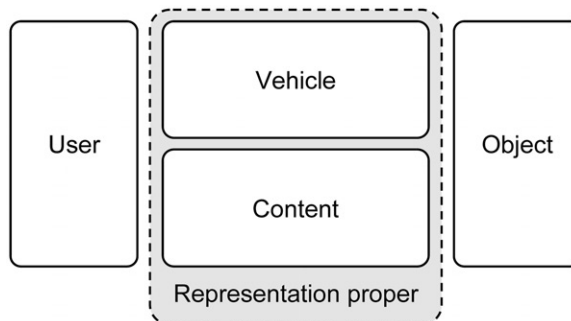


Figure 1 The basic elements of a representation.

understanding of the functionality of the mirror neuron system, one has to specify the user of the information that mirror neurons are supposed to carry. Yet, in most models on the working of mirror neurons the user remains unspecified. We will therefore analyze mirror neuron representations using just the vehicle, content and object aspect.

Single-cell recording experiments are based on what is called a *vehicle-first approach* (Uithol et al., 2008): one starts with a vehicle, in this case a neuron, and then tries to identify the type of stimuli the vehicle covaries with (i.e., the neuron responds to), thereby establishing a characterization of its content. Not all mirror neurons are equally selective in their responses. This has led Gallese et al. (1996) to discriminate three categories of mirror neurons: *strictly congruent*, *broadly congruent* and *non-congruent* mirror neurons. Strictly congruent mirror neurons respond to observed and executed movements that correspond both in terms of general action (e.g., grasping) and in the way that action was executed (e.g., precision grip). During action observation, the *object* of the representation is the movement of the experimenter or of another monkey. During action execution, the *object* is the movement of the monkey. The *content* of the neuron’s firing is assumed to be the shared feature of the two events that the neuron responds to, in this case the particular action with a particular grip (e.g., a grasp with a precision grip). When the motor and perceptual *object* share a common feature that gets reflected in the activity of the *vehicle*, the neuron is said to “mirror.”

3. Raising Levels of Abstraction

Each type of broadly congruent mirror neurons responds to a variety of grips or actions (Gallese et al., 1996), and consequently no commonality in the response profile can be found *at the level of grips*. For instance, broadly congruent neurons of

Table 1 The various mirror neurons, their response profile and their lowest common property in motor and visual response (i.e., their attributed content).

Type of mirror neuron	Response profile (M = Motor, V = Visual)	Lowest common property in motor and visual response profile
Non-congruent	M: Various actions V: Various actions	Object-related actions
Broadly congruent group 3	M: Specific action V: Various actions	Specific goals (grasping to eat)
Broadly congruent group 2	M: Specific hand action V: various hand actions	Specific category of actions (e.g. hand actions)
Broadly congruent group 1	M: Specific grip V: Various grips	Specific action (e.g. grasping with a hand)
Strictly congruent	M: Specific grip V: Specific grip	Specific grip (e.g. grasping with precision grip)

group 1 are highly specific to motor activity in terms of action and specific type of grip (e.g., a precision grip), but respond to the observation of various types of grips (e.g., both a precision grip and a full hand grip). See Table 1 for the various types of broadly congruent mirror neurons, their response profiles, and the lowest common property in the motor and visual response profile. Although it is not possible to specify a shared property on the level of grips, congruence *can* be found one level up, at the level of actions. Here the response profile is equally specific on the motor and perception side. The key property that mirror neurons owe their name to—the fact that the common property of a motor and a perceptual event gets reflected in the activity of one vehicle—can be preserved, but only by moving the description of the shared property from the level of grips up to the level of actions. The representational content attributed to this neuron can then be formulated as “grasping with the hand.”

In a similar vein, broadly congruent mirror neurons of group 2 can be taken to be congruent (and thereby representational) on the level of *categories of actions*, for instance hand actions versus non-hand actions. Neurons of group 3, in turn, can be considered congruent on the level of *action goals*, as these neurons appear to respond to the goal of an action and to be indifferent to the means by which this goal is achieved.

Non-congruent mirror neurons seem to show no clear-cut congruency between the observed action and the movement of the monkey. Hence, at first sight no common property seems available in their response profile. However, when the level of abstraction is raised to the level of object-directed versus non-object-directed actions, this neuron can be considered congruent and representational again, as mirror neurons only respond to *target-related* actions, and not to, for instance, mimed actions. The representational content of this type of neurons can thus be characterized as “*object-related actions*.” So even non-congruent mirror neurons can be made congruent by choosing the appropriate level of description.

In sum, neurons can be made to mirror—in the sense of representing a common property of a motor and perceptual event—by invoking levels of description of an increasing abstraction. As there exists an almost unlimited number of levels of abstraction, representational content can be attributed to *any* neuron responding to both executed and perceived actions. It must be emphasized that this is not a problem only to broadly congruent or non-congruent mirror neurons. The same interpretational principles can render a neuron congruent or incongruent anywhere along the continuum. This problem of unbounded content attribution undermines the explanatory value of the notion of mirroring,

By raising the level of abstractness, one strays from the rock bottom connotation of mirroring, making it increasingly difficult to see how highly abstract properties can be “reflected directly.” This problem can be overcome by imposing some principled restrictions on the level of abstraction at which mirroring can rightfully be said to occur. An analysis of the processes that are attributed to the MNS can offer such principled restrictions.

4. Action Understanding in the Mirror Neuron System

The human MNS is assumed to consist of the rostral part of the inferior parietal lobule, the lower part of the precentral gyrus and the posterior part of the inferior frontal gyrus (Rizzolatti & Craighero, 2004), although recently, neurons with mirror characteristics have been found outside these areas (Mukamel et al., 2010), in medial frontal and temporal cortices. This mirroring system is supposed to facilitate action understanding, goal understanding, and imitation by means of a mirroring process (Iacoboni et al., 2005; Iacoboni et al., 1999; Rizzolatti et al., 2001). Although the nature of the mirroring process is still largely unknown, some claims about features of this process can be found in the literature. The general idea is that perceptual representations of the observed action are mapped to motor representations of the observer's own action repertoire.³ Importantly, the process is assumed to be *direct* (i.e., the mirror neuron representation is brought about without involvement of higher, inferential processes, but by means of direct coupling, direct activation, direct association), or otherwise computationally *simple*. For example, Rizzolatti and Craighero write: "the proposed mechanism is rather simple. Each time an individual sees an action done by another individual, neurons that represent that action are activated in the observer's premotor cortex. . . . Thus, the mirror system transforms visual information into knowledge" (2004, p. 172). Similarly, Iacoboni writes: "we do not have to draw complex inferences or run complicated algorithms. Instead, we use mirror neurons" (2008, p. 7).⁴

Despite the general agreement on the simplicity of the mirroring process, there is diversity in the field when it comes to the capacity for abstraction of the mirroring process. It is claimed that the mirroring process produces representations of actions and action means (Buccino et al., 2004a; de Vignemont & Haggard, 2008; Fadiga, Craighero, & Olivier, 2005; Rizzolatti & Craighero, 2004), but at other places the scope of possible MNS output has been expanded to incorporate representations of the intentions behind actions (Gallese, Keysers, & Rizzolatti, 2004; Iacoboni, 2008; Iacoboni et al., 2005). For example, Rizzolatti & Sinigaglia claim that "through matching the goal of the observed motor act with a motor act that has the same goal, the observer is able to understand what the agent is doing" (2010, p. 269).

The range in abstraction attributed to the output of the mirror neuron system exceeds the one depicted in Table 1 for individual mirror neurons (from action means to immediate action goals). We will argue that, on theoretical grounds, it is implausible that a direct or otherwise simple process has the capacity for reliably producing representational content at or above the level of action goals. In order to do so we will have to make minimal assumptions on what could be possibly meant by "processing" in the context of mirroring. For our purposes it will suffice to assume that by processing one means a form of "computation," in the broad sense of the word (Chalmers, 1995; Eliasmith, 2010; Piccinini, 2006). This would include non-traditional and non-symbolic forms of computation—such as the various forms of neural network computations—but it would exclude hypothetical mechanisms with

presumed computational processing powers that have no possible physical implementation (see also Frixione, 2001; Tsotsos, 1990; and van Rooij, 2008).

5. Goals Inference is Context Dependent

The recognition of an action alone is not sufficient for a reliable goal inference, as multiple goals can be achieved by a given action (e.g., picking up a cup for drinking, pouring, cleaning up; de Vignemont & Haggard, 2008; Jacob & Jeannerod, 2005). Also, multiple actions can be performed to achieve a certain goal (the goal to drink with grasping a cup, ordering a beer, opening the tap). In other words, there is not a one-to-one mapping between actions and goals, but a many-to-many mapping. Therefore, goal-action associations alone cannot produce a unique goal when observing a given action.

Which goal can be reached by an action is dependent on the *context* in which the action is performed. For instance, hand waving can be a means to shooing away mosquitos as well as making a taxi stop. It is the context of the action—the presence of taxis or mosquitoes, an urban environment or a camp ground—that leads to a different interpretation of the observed action. So in order to reliably infer goals from observed actions, both the action and the context must be processed (De Ruiter, Noordzij, Newman-Norlund, Hagoort, & Toni, 2007; Jacob & Jeannerod, 2005; Kilner, Friston, & Frith, 2007b; Toni, Lange, Noordzij, & Hagoort, 2008; van Rooij, Haselager, & Bekkering, 2008). This context-dependency of goals is also present in Iacoboni and colleagues' tea-cup experiment (Iacoboni et al., 2005), where grasping a cup from a neatly set table was used to suggest the goal of “drinking,” and grasping it from a messy table to suggest “cleaning up.” Typical mirror areas (the posterior part of the inferior frontal gyrus and the adjacent sector of the ventral premotor cortex) were shown to be more active when an intention could be inferred from one of these contexts than in cases of no context, which the authors took as evidence that a mirroring process is responsible for performing the context-dependent inferring of goals (Iacoboni, 2008; Iacoboni et al., 2005). We argue, however, that although it might be the case that parts of the inferior frontal cortex are involved in context-dependent goal inference, it does not appear likely that this is done by means of a mirroring mechanism. The degree of context understanding required for reliable goal inferences seems, given the computational complexity of this task, to exceed by far the abilities of a direct or otherwise simple mechanism.

6. Goal Inference is “Non-Direct” and “Complex”

Goal inference above the level of immediate goals cannot be based on a *direct* or a *simple* mechanism. Although it remains somewhat implicit what most researchers mean by direct, we will argue that it cannot mean that sensory representations are mapped on the motor system without significant aid of other, more inferential processes. Such a *direct* mapping mechanism would need access to a mapping

structure in which each possible action-context combination maps directly to a unique goal. But given the number of possible actions and the number of possible relevant context aspects, this strategy would, due to what is called a combinatorial explosion, very soon result in an unmanageable number of mappings. To illustrate, consider there are, say, just 35 possible context features—e.g., that the person in a white coat is a surgeon (or a psychopath); the scalpel is sterile (or poisoned); the setting is a hospital (or a movie set); the person being cut is a patient (or an actor); the other people in the room are nurses (or medical students, etc.)—and there are several different goals (e.g., grasping the scalpel to cure, to hurt, to cut wire, to clean, to put away, to give to a nurse, etc.).⁵ If we make the simplifying assumption that any context feature is either “present” or “absent” then there already exist more than 100,000,000,000 distinct possible contexts (that is, one hundred billion, roughly the number of neurons in the entire human brain). Allowing values in between “present” and “absent” only serves to increase the number more. This unmanageable number of combinations makes it impossible to solve goal inference using a direct mapping solution. In everyday conditions, people can easily infer goals from observed actions in a certain context. This suggests that people do not instantiate all possible action-context combinations. Instead, it suggests that action to goal mappings must be “non-direct” or “inferential,” in the sense that they use some form of knowledge about the *interaction* between actions and contexts to infer plausible goals.

Besides statements about mirroring being a direct process, claims that context-sensitive goal inferences could be achieved by a *simple* mechanism can also be found in the literature (Iacoboni, 2008; Rizzolatti & Craighero, 2004). There are, however, reasons to believe that *no* simple mechanism can support a general capacity for making context-sensitive goal inferences, as these inferences belong to a class of inferences that are known to be notoriously difficult. Inferring goals from actions is a form of *abduction*, also called *inference to the best explanation* (Baker, Goodman, & Tenenbaum, 2008; Charniak & Goldman, 1993; Haselager, 1997). In abduction, causes are hypothesized to explain observed effects. In the case of goal inference, the cause is a “goal” and the observed effect is an “action.” Existing models of abduction of a reasonable generality all belong to the class of so-called computationally intractable (or *NP-hard*) functions (Abdelbar & Hedetniemi, 1998; Bylander, Allemang, Tanner, & Josephson, 1991; Eiter & Gottlob, 1995; Garey & Johnson, 1979; Nordh & Zanuttini, 2005, 2008; Thagard & Verbeurgt, 1998; van Rooij, 2008). These are functions strongly conjectured by mathematicians to defy efficient computation by any physically implementable mechanism (see Garey & Johnson, 1979; and van Rooij, 2008, for details). This thus suggests that it is unlikely that a reasonably general capacity for making goal inferences is based solely on a mechanism that qualifies as “simple.”

7. Taking Complexity into Account

Due to the tractability issues with abduction in its general form, it seems unlikely that humans can perform completely domain general goal inferences. Instead, when

trying to figure out a goal behind an action, humans may be performing this task against the background of a restricted domain of situations. This restricted domain must still be quite general if it is to account for the variety of situations in which humans can infer goals. At the same time it needs to be sufficiently constrained in nature to allow for tractable goal inference. At present it is unclear how tractable models of abduction can be formulated without rendering their domain of application too simple for modeling real world domains (see, e.g., Nordh & Zanuttini, 2005, 2008).

It is noteworthy that computational models of goal inference in cognitive science that seem to work (i.e., that make plausible goal inferences without running into tractability issues) severely restrict the possible contexts and the number of possible actions and goals, keeping their application domain far removed from realistically complex situations (Baker, Tenenbaum, & Saxe, 2007; Cuijpers, Van Schie, Koppen, Erlhagen, & Bekkering 2006; Erlhagen, Mukovskiy, & Bicho, 2006; Oztop, Wolpert, & Kawato, 2005). For example, Baker et al. (2007) modeled goal inferences made by an observer viewing a point moving in the flat plane to one of three possible goal states, and Oztop et al. (2005) modeled goal inferences made by an observer viewing a reaching movement in the flat plane to one of eight possible goal states, and a grasping movement with three possible goals. The availability of such successfully predictive, albeit highly restricted, models, seems to lead to an underestimation of the computational complexity inherent in more general domains.

A possible way of dealing with the limitations of a direct-matching mechanism and the fact that it cannot account for goal inference above the level of immediate goals could be to “upgrade” the notion of mirroring, by supplying it with more inferential capabilities. This upgraded process might thereby be capable of explaining a substantial part of goal inference after all. A problem with such an approach, however, is that calling this complex, unknown form of processing a “mirroring process” does not provide additional explanatory value over calling it “processing.” Moreover, it further increases the risk of underestimating the complexity of the computational problem underlying goal inference.

Once we recognize the difficulties inherent in making context-sensitive goal inferences, we can better address the question of how “mirroring” can contribute to an explanation of goal inference. To draw a parallel: line detection is likely to be an important sub-process of visual perception, yet we do not think of object recognition as a case of mere line detection. Likewise, mirroring may be an important sub-process in goal inference, but we would not do well to assume that goal inference is a case of mere mirroring. In this light it is informative that computational models that set out to connect to neuroanatomy place the major burden of processing context in goal inferences outside the MNS (Kilner, Friston, & Frith, 2007a; Kilner et al., 2007b; Oztop, Kawato, & Arbib, 2006). This separation between context-processing and mirror neuron processing makes sense given that context is such a multimodal and multifaceted construct; it is unlikely that a mirroring process is by itself capable of processing context in all its complexity.

These considerations can help to interpret the firing characteristics of mirror neurons in the sense that when the attributed content becomes increasingly abstract, it becomes less plausible that the mirror neuron is the actual functional unit that is playing the key role in the cognitive function one is interested in.⁶ Instead, at higher levels of abstraction, such as higher level action goals or intentions, it becomes increasingly plausible that less direct (more inferential) cognitive processes influence the response profile of mirror neurons.

An intuitive response to our analysis is that we do not take the role of expectations into account. In our daily lives we are familiar with many different contexts and appropriate actions in that context. Such associations could result in an expectation or prediction of what the goal of an observed action is, which can help goal inference on many occasions. Indeed, most of the time a reaching action has grasping as its goal, and most grasped cups are grasped for drinking. But it is important to realize that such default associations can only predict one goal given a certain action or context-action pair. People can easily ignore such default interpretations when necessary (say, when the waiter is grasping your empty cup). It is precisely this capacity to sometimes use a default association and sometimes use another context sensitive process that makes direct matching an insufficient mechanism to account for our general capacity to infer goals from actions.

Another objection to our arguments could be that there is data available that does show mirror neurons representing higher intentions. For instance, Fogassi et al. (2005) found mirror neurons in the monkey's inferior parietal lobule that responded selectively for different intentions underlying the same actions. Monkeys were trained to grasp a piece of food and either place it in a container on their shoulder, or eat it. Some neurons responded differently for these two intentions. Importantly, in some neurons this difference in firing was preserved when the monkeys observed the experimenters perform the same actions. Fogassi et al., take this as evidence that mirror neurons are in fact capable of mirroring intentions.

At a first sight this data seem in contradiction with our analysis, as we argued that a mirror mechanism cannot yield an action description at the level of intentions. However, as it is pointed out by (among others) Csibra (2007), from the fact that this activity is indicative of intention understanding one cannot conclude that it is constitutive of intention understanding. A mechanism in which low-level mirror processes are modulated by other, context sensitive processes could result in similar activity in these neurons.

Notwithstanding these interpretational difficulties, these data and, for instance, Umiltà et al.'s (2008) finding of goal-sensitive motor neurons⁷ led Rizzolatti and his colleagues to reformulate their direct-matching hypothesis. According to this adapted hypothesis, mirroring is a dual-route process with one route directly matching movements and the other mapping the goal of the observed motor act onto the observer's own motor repertoire (Rizzolatti & Sinigaglia, 2010). This proposal circumvents the problem of directly mapping actions to goals, thereby sidestepping the tractability issues described above. However, it does not offer an alternative explanation of how the observed goals are recognized or processed.

We do not want to deny the importance of mirroring to action recognition and goal recognition, but we want to stress that mirroring, conceived as direct-matching, is unlikely to be the mechanism that facilitates goal attribution in all its complexity. The activation of areas outside of the classical mirror neuron areas upon interpreting actions found in imaging studies, such as occipital, posterior parietal, parietofrontal, and prefrontal areas (Calvo-Merino, Glaser, Grezes, Passingham, & Haggard, 2005; Evangelidou et al., 2009; Grèzes & Decety, 2001; Iacoboni et al., 2005) hints to a modulating role of these brain systems. We think that, in addition to characterizing the nature of the mirror motor representations (de Vignemont & Haggard, 2008), unraveling this interaction between different brain parts is a necessary step in addressing the problem of how people are capable of understanding observed actions and inferring the goals those actions serve. Subsuming complex processes under the label of “mirroring” does not do justice to the complexity of the process of inferring goals from observed actions.

8. Conclusion

Research based on single-cell recordings is hindered by a conceptual indetermination that allows for content attribution of an unbounded abstraction, troubling the interpretation of the data of these experiments. However, research into the MNS as a whole, together with an analysis of the computational complexity of the attributed task can help to restrict the class events that can be said to be mirrored. We have argued that, on theoretical grounds, a direct or otherwise simple (mirror-like) process cannot be used to infer action goals, as the context-dependency of goals defies a simple or direct solution to the task of goal inference. We therefore propose to restrict the use of the term mirroring to describe a simple reflexive mechanism that is involved in relatively low-level action observation and recognition, such as grips of basic actions. This restriction can help interpreting the data acquired by means of single-cell recordings.

It is, of course, obvious that we humans are very well capable of deriving intentions from observed actions, in spite of the presumed complexity of this task. Explaining how we solve or evade the apparent computational intractability inherent in context-dependent goal inference is a major question for cognitive neuroscience. However, we think that answering this question is not helped by heaping together potentially complex processes under the label of “mirroring.” Instead, much work in various areas of cognitive science needs to be done before this question can be answered. How good, exactly, are we at goal inference? In what circumstances do we make mistakes? In what way is context relevant for goal attribution? What restrictions apply to the action domain that make goal inference computationally tractable enough for humans in everyday life? Answers to these questions can guide future theory development on goal inference. It is only when the complexity of the task is appreciated to the full extent that we can expect to get insight into how goal inference is achieved by brain mechanisms and how mirroring contributes to these mechanisms.

Notes

- [1] Technically, a consequence of this inability to measure individual neurons in humans is that mirror neurons have not yet been unequivocally established in humans. There is indirect evidence of mirror neurons in humans based on repetition suppression (Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010), but this result is not unequivocal (Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Lingnau, Gesierich, & Caramazza, 2009).
- [2] This contrast is also denoted with the terms ‘intention in action’ and ‘prior intention’ (Searle, 1983).
- [3] The activity of the mirror neuron system is often described as a form of resonance. This resonance is claimed to be either *interpersonal*, i.e., between parts of the premotor system of the observer and of the executor; or *intrapersonal*, i.e., between a visual and a motor representation in the observer. See for an elaboration on this distinction (Uithol, van Rooij, Bekkering, & Haselager, forthcoming).
- [4] Strictly speaking, this statement reflects a category mistake in the sense that “we” can use mirror neurons. One may wonder what the “we” consists of if neurons are not part of it. However, we do not want to elaborate on this along the lines of Bennett and Hacker (2003). Rather, we see this statement (and many similar others) as a “rough and ready” type of description that could be formulated more appropriately (e.g., by speaking of mirror neurons that implement our “capacity to...”) when the occasion requires it.
- [5] Evidently, real world situations can vary in many more than just 35 features, but to make our case we do not need to assume any more such features.
- [6] These limitations, however, are as such not enough to provide a solid and undisputable bound for content attribution to the firing of individual mirror neurons. That would require a deep understanding of the functioning of the MNS at the neuronal level, as well as taking an ontological position on the nature of representation, which would be beyond the scope of this article.
- [7] Umiltà et al. (2008) found motor neurons that fired selectively for pinching with pliers, irrespective of whether the pinching was achieved with normal or reverse pliers (so irrespective of whether the hand had to close or open in order to pinch). While we agree that this is indeed an abstraction from the basic kinematics of a motor act, it is still a relatively basic act, remote from, say, Iacoboni et al.’s, “intention to clean the table.” Moreover, the same reserve as presented in the discussion of Fogassi et al.’s results is relevant here: from the data we cannot conclude whether the firing is indicative or constitutive of goal understanding. In addition, as this study analyzed motor neurons, not mirror neurons, the relation to a mirroring process in goal recognition is not elucidated.

References

- Abdelbar, A. M., & Hedetniemi, S. M. (1998). Approximating MAPs for belief networks is NP-hard and other theorems. *Artificial Intelligence*, 102, 21–38.
- Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based social goal inference. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Washington, 1447–1452.
- Baker, C. L., Tenenbaum, J. B., & Saxe, R. R. (2007). Goal inference as inverse planning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*, Austin, TX, 779–784.
- Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist’s challenge in cognitive science. *Cognitive Science*, 22, 295–318.

- Bekkering, H., & Wohlschläger, A. (2002). Action perception and imitation: A tutorial. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Vol. 19. Attention and performance*. Oxford: Oxford University Press.
- Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Malden, MA: Blackwell.
- Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem?. *Trends in Cognitive Sciences*, 9, 489–495.
- Buccino, G., Binkofski, F., & Riggio, L. (2004a). The mirror neuron system and action recognition. *Brain and Language*, 89, 370–376.
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., et al. (2004b). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, 42, 323–334.
- Bylander, T., Allemang, D., Tanner, M. C., & Josephson, J. R. (1991). The computational-complexity of abduction. *Artificial Intelligence*, 49, 25–60.
- Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15, 1243–1249.
- Chalmers, D. J. (1995). On implementing a computation. *Minds and Machines*, 4, 391–402.
- Charniak, E., & Goldman, R. P. (1993). A Bayesian model of plan recognition. *Artificial Intelligence*, 64, 53–79.
- Chong, T. T., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology*, 18, 1576–1580.
- Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition: Vol. 12. Attention and performance*. Oxford: Oxford University Press.
- Cuijpers, R., Schie, H. T., Koppen, M., Erlhagen, W., & Bekkering, H. (2006). Goals and means in action observation: A computational approach. *Neural Networks*, 19, 311–322.
- Cummins, R. (1989). *Meaning and mental representation*. Cambridge, MA: MIT Press.
- De Ruyter, J. P., Noordzij, M., Newman-Norlund, S., Hagoort, P., & Toni, I. (2007). On the origin of intentions. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition: Vol. 12. Attention and performance*. Oxford: Oxford University Press.
- de Vignemont, F., & Haggard, P. (2008). Action observation and execution: What is shared? *Social Neuroscience*, 3, 421–433.
- de Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences*, 10, 435–441.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, 91, 176–180.
- Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). A mirror up to nature. *Current Biology*, 18, R13–R18.
- Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Eiter, T., & Gottlob, G. (1995). The complexity of logic-based abduction. *Journal of the Association for Computing Machinery*, 42, 3–42.
- Eliasmith, C. (2010). How we ought to describe computation in the brain. *Studies in the History and Philosophy of Science*, 41, 313–320.
- Erlhagen, W., Mukovskiy, A., & Bicho, E. (2006). A dynamic model for action describe and goal-directed imitation. *Brain Research*, 1083, 174–188.
- Evangelidou, M. N., Raos, V., Galletti, C., & Savaki, H. E. (2009). Functional imaging of the parietal cortex during action execution and observation. *Cerebral Cortex*, 19, 624–639.
- Fadiga, L., Craighero, L., & Olivier, E. (2005). Human motor cortex excitability during the perception of others' action. *Current Opinion In Neurobiology*, 15, 213–218.

- Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *The European Journal Of Neuroscience*, *17*, 1703–1714.
- Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: Mirror neurons for execution, observation, and imagery. *Neuroimage*, *37*, 1315–1328.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*, 662–666.
- Frixione, M. (2001). Tractable competence. *Minds and Machines*, *11*, 379–397.
- Gallese, V. (2001). The 'shared manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies*, *8*, 33–50.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–610.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*, 493–501.
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396–403.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts. *Cognitive Neuropsychology*, *22*, 455–479.
- Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco: Freeman.
- Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.
- Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*, 590–616.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*, 1–19.
- Haselager, W. F. (1997). *Cognitive science and folk psychology: The right frame of mind*. London: Sage.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *The Behavioral and Brain Sciences*, *24*, 849–877.
- Iacoboni, M. (2008). *Mirroring people: The new science of how we connect with others*. New York: Farrar, Straus and Giroux.
- Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*, 942–951.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*, e79.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.
- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind & Language*, *23*, 190–223.
- Jacob, P., & 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. *The Behavioral and Brain Sciences*, *17*, 187–201.
- Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Progress In Brain Research*, *156*, 379–402.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007a). The mirror-neuron system: A Bayesian perspective. *NeuroReport*, *18*, 619–623.
- Kilner, J. M., Friston, K. J., & Frith, C. D. (2007b). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*, 159–166.
- Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, *29*, 10153–10159.

- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9925–9930.
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron responses in humans during execution and observation of actions. *Current Biology*, *20*, 750–756.
- Nakahara, K., & Miyashita, Y. (2005). Understanding intentions: Through the looking glass. *Science*, *308*, 644–645.
- Newman-Norlund, R. D., Van Schie, H. T., Van Zuijlen, A. M. J., & Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience*, *10*, 817–818.
- Nordh, G., & Zanuttini, B. (2005). Propositional abduction is almost always hard. *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, Edinburgh, UK, 534–539.
- Nordh, G., & Zanuttini, B. (2008). What makes propositional abduction tractable? *Artificial Intelligence*, *172*, 1245–1284.
- Oztop, E., Kawato, M., & Arbib, M. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, *19*, 254–271.
- Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, *22*, 129–151.
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*, 179–217.
- Piccinini, G. (2006). Computation without representation. *Philosophical Studies*, *137*, 205–241.
- Prinz, W. (1997). Perception and action planning. *The European Journal Of Cognitive Psychology*, *9*, 129–154.
- Rizzolatti, G. (2005). The mirror neuron system and imitation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation* (pp. 55–76). Cambridge: MIT Press.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review Of Neuroscience*, *27*, 169–192.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–142.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–670.
- Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*, 264–274.
- Saxe, R. (2005). Tuning forks in the mind. *Trends in Cognitive Sciences*, *9*, 321.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.
- Shea, N. (2007). Content and its vehicles in connectionist systems. *Mind & Language*, *22*, 246–269.
- Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, *22*, 1–24.
- Toni, I., Lange, F. P., Noordzij, M. L., & Hagoort, P. (2008). Language beyond action. *Journal of Physiology*, *102*, 71–79.
- Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *The Behavioral and Brain Sciences*, *13*, 423–469.
- Uithol, S., Haselager, W. F., & Bekkering, H. (2008). *When do we stop calling them mirror neurons?* In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society, Austin, TX*, 1783–1788.
- Uithol, S., van Rooij, I., Bekkering, H., & Haselager, W. F. G. (forthcoming). Understanding motor resonance. *Social Neuroscience*.
- Umiltà, M. A., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., et al. (2008). When pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences*, *105*, 2209–2213.

- van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, 32, 939–984.
- van Rooij, I., Haselager, W. F., & Bekkering, H. (2008). Goals are not implied by actions, but inferred from actions and contexts. *The Behavioral and Brain Sciences*, 31, 38–39.
- Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40, 655–664.
- Wohlschläger, A., & Bekkering, H. (2002). Is human imitation based on a mirror-neuron system? Some behavioural evidence. *Experimental Brain Research*, 143, 335–341.