The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/73327

Please be advised that this information was generated on 2018-10-11 and may be subject to change.
When Do We Stop Calling Them Mirror Neurons?

Sebo Uithol (S.Uithol@nici.ru.nl)
Willem F. G. Haselager (W.Haselager@nici.ru.nl)
Harold Bekkering (H.Bekkering@nici.ru.nl)
Nijmegen Institute for Cognition and Information
Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Abstract

The discovery of mirror neurons in the 1990s has led to much excitement in the cognitive neurosciences. After the initial discovery more and more abilities have been attributed to these neurons. As mirror neurons are commonly viewed as vehicles of representation, we analyze the increasingly wider representational role mirror neurons play and argue for a principled distinction between mirror and non-mirror neurons.

Keywords: mirror neurons; representation; action recognition; goal understanding.

Introduction

In 1992 Di Pellegrino and his colleagues discovered that neurons in the rostral part of the inferior premotor cortex of the macaque brain fire both during the execution and the observation of an action (1992). Because of the double representational role these neurons play, they were later dubbed ‘mirror neurons’ (Gallese et al., 1996; Rizzolatti et al., 1996). The discovery of mirror neurons caused great excitement in the cognitive neurosciences, as these neurons seem to offer a solid, neuronal base for the coupling of perception to action. Over time, more and more abilities were attributed to these neurons. Mirror neurons were deemed to be involved in the inference of intentions and goals (Fogassi et al., 2005; Iacoboni et al., 2007). Also mirror properties were connected to other modalities such as hearing (Keysers et al., 2003) and touch (Keysers et al., 2004) and found in other brain regions (Gallese et al., 2002).

The attribution of increasingly general abilities to mirror neurons has diminished the original clarity on what mirror neurons are, what they do and how they could be capable of performing the functions attributed to them. For example, a debate has risen about whether mirror neurons show that direct matching of low-level motor activity is sufficient for describing the coupling of perception to action (Rizzolatti & Craighero, 2004), or that it should still be accompanied by the goal-directed hypothesis (Erlhagen, Mukovskiy & Bicho, 2006; Koski et al., 2002). Also, the use of mirror neurons as support for the simulation theory of mind reading (Gallese & Goldman, 1998) has been questioned (Csibra, 2005, 2007; Saxe, 2005).

An analysis of the characterizing properties of mirror neurons and the representational roles they may and may not play can prove helpful to distinguish between neurons that could rightfully be called ‘mirror neurons’ and neurons that have other – nevertheless interesting – properties.

Mirror Neurons as Representations

When one wants to emphasize the special character of mirror neurons, one naturally adopts a representational point of view. For example, when Gallese et al. (1996, p. 606) speculate upon the role of mirror neurons, they state that “[a]nother possible function of mirror neuron movement representation is that this representation is involved in the ‘understanding’ of motor events”. Likewise, Rizzolatti et al. (1996, p. 131) propose that “that [mirror neuron’s] activity ‘represents’ the observed action.” The idea that mirror neurons carry representational content is based on covariance as measured, in particular by single cell recordings. Every time the monkey executes a particular movement or observes that particular movement being executed by the experimenter, a neuron fires.

Representations based on covariance are ubiquitous in daily life. For instance, we regard the gas meter a representation of the amount of fuel in our tank precisely because there is a reliable covariance between the angle of the meter and the level of fuel in the tank. There have been arguments showing that a reliable covariance is neither a sufficient (Haugeland, 1991) nor a necessary (Millikan, 1984) condition for representation, but we want to jump over these foundational difficulties, as we do not want to argue for or against the representational view in general (see e.g. Beer (2000), Clark (1997), Haselager, De Groot & Van Rappard, (2003) and Markman & Dietrich (2000) for various positions in this debate).

A representation consists of a vehicle and a content and relates to a user and an object. These elements have been visualized in figure 1. 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 or the content can be of a more general or more abstract nature than the object represented (e.g. “a sparrow” can get represented as “a bird”). 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

1783
behavior. As the user is mostly unspecified in case of mirror neuron representations, we will pay little attention to this aspect of representation here.

During action execution, the experimenter or the movement of another monkey. Of the strictly congruent mirror neuron is the movement of the monkey. In both cases the representation is the particular action (the vehicle).

During action observation, the object of the representation of the strictly congruent mirror neuron is the movement of the experimenter or the movement of another monkey. During action execution, the object of this representation is the movement of the monkey. In both cases the content of the representation is the particular action (the means towards an end, for example grasping through a precision grip). The content is abstracted from the performer of the movement, as the neurons fire equally in response to movements made by the monkey or the experimenter, so no information on the executor is included here. The neuron itself is the vehicle of the representation.

We can now reformulate what is special about this type of neurons: one neuronal vehicle covaries its activities with two objects or events of apparently different domains (action and perception) that share the same representational content. Neurons ‘mirror’ when different objects share a common property that gets reflected in the activity of the vehicle.

**Representational Content and Levels of Abstraction**

With the *broadly congruent* mirror neurons things are not as straightforward. These neurons display a connection, but not identity, between the observed and executed action and appear to be more specific on the motor side than on the perceptual side. Gallese et al. (1996) discern three groups of broadly congruent mirror neurons. Neurons of group 1 are highly specific for motor activity in terms of action and specific type of grip, but respond to the observation of various types of grips. An example of a neuron of this group is a neuron that fires only when the monkey grasps an object using a precision grip, and not with any other type of grip, but also when the experimenter grasps the object with various kinds of grips, unlike strictly congruent neurons, that fire only at the observation of one specific grip type. So, when speaking at the level of grips, it is not possible to specify the shared property, and hence the representational content of these neurons cannot be formulated. However, congruence can be found one level up, i.e. the level of actions, because from this perspective the response profile is equally specific on the motor and perception side, namely actions, e.g. grasping. The key property that mirror neurons of the strictly congruent type owe their name to – the fact that the common property of two different events 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 on to the level of actions.

Neurons of group 2 become active during one motor action with a hand, but visually respond to two or more different hand actions. Like the neurons of group 1, the content is more detailed on the motor side than on the observation side. Here the congruence is only preserved for categories of actions. This demands climbing yet another level of generality, i.e. by distinguishing hand actions and non-hand actions. So the representational content of these neurons can only be described at this level of broad action categories.

The activation of neurons of group 3 is dependent on the goal of an action, regardless of how it was achieved. These action-dependent neurons are neither specific on the motor side, nor on the perception side. The level on which congruence can be found is even higher than that of neurons of group 2, as hand actions (for example grasping as well as other actions (grasping with the mouth) can serve a common goal (grasping to eat)).

According to Gallese et al. (1996), mirror neurons of the non-congruent group exhibit no clear-cut relationship between the observed action and the movement of the monkey. So their claim is that the response profile of these neurons is too different on the perceptual and motor side to find a level of analysis in which a shared property can be found. At first sight, this makes it impossible to specify or characterize the representational content of these individual neurons, as it is difficult to find an informative description of the property on motor and perception side that the neuron covaries its activity with.
Table 1: The various mirror neurons arranged according to their congruence at different levels of abstraction.

<table>
<thead>
<tr>
<th>Type of mirror neuron</th>
<th>Response profile (M=Motor, V=Visual)</th>
<th>Lowest common property in motor and visual response profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non Congruent</td>
<td>M: Various actions</td>
<td>Object-related actions</td>
</tr>
<tr>
<td></td>
<td>V: Various actions</td>
<td></td>
</tr>
<tr>
<td>Broadly Congruent group 3</td>
<td>M: Specific action</td>
<td>Specific goals (grasping to eating)</td>
</tr>
<tr>
<td></td>
<td>V: Various actions</td>
<td></td>
</tr>
<tr>
<td>Broadly Congruent group 2</td>
<td>M: Specific hand action</td>
<td>Specific category of actions (e.g. hand actions)</td>
</tr>
<tr>
<td></td>
<td>V: various hand actions</td>
<td></td>
</tr>
<tr>
<td>Broadly Congruent group 1</td>
<td>M: Specific grip</td>
<td>Specific action (e.g. grasping with a hand)</td>
</tr>
<tr>
<td></td>
<td>V: Various grips</td>
<td></td>
</tr>
<tr>
<td>Strictly Congruent</td>
<td>M: Specific grip</td>
<td>Specific grip (e.g. grasping with precision grip)</td>
</tr>
<tr>
<td></td>
<td>V: Specific grip</td>
<td></td>
</tr>
</tbody>
</table>

Even here however, a higher level can be found to formulate the shared property of the events involved, namely the fact that the actions involve objects. Movements mimicking an object related action evoke no response so the shared property that gets reflected in the activity of the neuron could be labeled as “object-related actions”.

More levels than the ones mentioned by Gallese et al. could be postulated. Theoretically, below the level of grips there can be thought to be a level of precise motor execution. At this level no shared property can be found that gets reflected in the activity of the neurons, as even the most fine-grained mirror neurons – the strictly congruent mirror neurons – allow for small variations in the execution of an action.

The types of neurons and their lowest common property in motor and visual response profile are shown in Table 1. The tendency is obvious: when a shared property cannot be specified at one level, one can go up one level and use a new idiom in which commonalities can be found. It is possible to formulate a level of abstraction on which all neurons are incongruent, but also a level can be formulated on which all neurons are congruent.

In all, neurons can be made to mirror, in the sense of reflecting a common property of two events, by invoking levels of description of an increasing abstractness. The representational analysis has shown that, when moving through the different categories of mirror neurons, the representational content becomes of an ever more general nature in order to be able to continue speaking of mirroring.

**Vehicle-First Approach**

The continuous search for higher levels of abstraction (from action recognition to goal understanding) in which the common factor in the observed and performed action can be expressed is possibly due to the focus on response profiles of individual neurons. This is a logical consequence of the original approach taken in the monkey experiments, i.e. single cell recordings. In this setup, one presumes a neuron to be a vehicle of content and looks for the objects or events that it might covary its activity with, whereupon a conclusion is drawn regarding the potential content the neuron’s firing might carry.

If one is looking for the representational content of a single neuron, one assumes that this single neuron is a vehicle by itself. This might be the case for some actions, as the discovery of strictly congruent mirror neurons seem to suggest, but certainly need not be the case for every action and every neuron. There is always a higher level at which there is a description of the behavior of the neurons possible, but at one point, one may start to wonder whether this continuously more abstract interpretation of what supposedly gets mirrored is fruitful or warranted.

Things will become even more problematic when not only local coding is considered, but also its contrasting coding scheme, distributed coding. In a local coding scheme the activity of a single neuron or group of neurons is sufficient for representing a certain property (see Van Gelder (1999) for an explanation of various types of coding). The proverbial “grandmother-neuron” is the most famous example of this type of coding. By contrast, in a distributed coding scheme an item is represented by the pattern of activity over a more than minimal extent of the resources available for representing. In this case the pattern of activity is the vehicle of the representation, not the individual nodes that the vehicle consists of. This type of coding is often discussed in the context of neural network models (e.g. Van Gelder (1992; 1999)). In distributed representation, there need not be a structure where a subpart of the vehicle represents a subpart of the content. In that case, the vehicle
as a whole carries the entire content and the subparts of the vehicle do not carry identifiable content by themselves. Locating a vehicle without knowledge of the content is highly problematic in a distributed, unstructured coding scheme.

To illustrate this, imagine that we have a neural network capable of representing various items. And suppose that it does so by exhibiting a unique pattern of activation over all the units in the network’s output layer. In this case the vehicle consists of multiple neurons. There need not be a structure in the sense that identifiable subparts of the content correspond to identifiable subparts of the vehicle (i.e. neurons), in which case a subpart of the vehicle makes a contribution to the entire content (again, see Van Gelder (1992; 1999) for elaboration on this topic). When we record one node in the layer we will not find unequivocal representational content because this node does not carry any straightforwardly identifiable content by itself.

With single cell recordings problems might be similar. When the interpretation of the representational content of a neuron’s activity reaches a level of abstraction of a less plausible height, this can suggest that the activity of the neuron is part of a distributed representation. In such cases a search for the common property to be reflected is neither necessary nor likely to produce illuminating results. At higher levels of action interpretation it seems more likely that neuronal systems rather than individual neurons are providing the basic processing elements, so the focus shifts from mirror neurons to the mirror neuron system.

With the widening of the scope, from mirror neurons to mirror neuron system, the attributed task grows accordingly. More competencies are attributed to the system while, at the same time, holding on to the characteristics of individual neurons. The problems involved in this strategy will be discussed in the next paragraph.

**From Action Recognition to Intention Understanding**

As research on mirror neurons and the mirror neuron system continued, ever more competencies were attributed to the neurons and the system. For instance, mirror neurons are generally appreciated as the solid neuronal basis for the *understanding* of actions (Nakahara & Miyashita, 2005; Rizzolatti & Craighero, 2004). Recognized specific actions are directly related to the motor system which facilitates the understanding of these actions or their underlying goals (Fogassi et al., 2005; Gallese et al., 1996). According to Rizzolatti et al. (2001) an action is “understood” when its observation causes the motor system of the observer to resonate. This resonance is supposed to lead to the same activity in the motor system as would be the case when the observer would perform the action, which facilitates the understanding of the observed action.

Iacoboni et al. (2005) claim that not only is the mirror neuron system involved in the understanding of actions, it is also used for the detection of intentions of others. They base their claim on the fact that activity in areas associated with the mirror neuron system is dependent on whether intention of an action can be inferred. When an intention has to be inferred from a context (cup grasping for drinking versus cup grasping for cleaning up), there is a significant increase in signal in the parieto-frontal cortical circuit for grasping.

We have argued that when the representational content of mirror neurons gets of an increased level of abstractness, the activity of these neurons can no longer be rightfully described as mirroring. As argued, the suggestion that mirror neurons are basic to action recognition is intuitively plausible when applied to the level of specific grips and relatively straightforward actions such as grasping, but it is quite another thing to suggest that the same basic neuronal mechanism underlies such a high level process as intention understanding. The goal of an action has to be inferred from the recognized action using context (van Rooij, Haselager, & Bekkering, in press), past experiences with the observed actor (Ferrari, Rozzi, & Fogassi, 2005), and a lot of background knowledge. That is an awful lot for a single neuron to directly mirror onto the motor system. There seems to be too much inference and knowledge consultation involved in goal understanding in order to plausibly characterize the underlying process as a case of pure mirroring. When climbing to higher levels of abstraction, the attributed function of the neurons shifts from resonating with something readily observable to making inferences about a hidden – or at least indirectly observable – feature. With every step up, the distance between the attributed content and the observed input becomes larger. This provides a reason to be skeptical, as the task of the individual neuron grows to a questionable size. A lot of processing has to take place before these neurons can be specific to a particular goal of an action. Also, it is far from obvious how a notion as abstract as a goal can be mapped directly to a motor system, causing the right resonance to occur.

Of course, Iacoboni et al. do not claim that a single neuron is capable of recognizing intentions. Hence they speak of a mirror neuron system instead of mirror neurons. And naturally, a system may be able to accomplish a task that its parts cannot accomplish by themselves. Yet, they depict the mirror neuron system as a system consisting of mirror neurons and they explain the working of the system by describing the working of mirror neurons.

[...] the intentions behind the actions of others can be recognized by the motor system using a mirror mechanism. Mirror neurons are thought to recognize the actions of others, by matching the observed action onto its motor counterpart coded by the same neurons. (p.533)

The findings of their fMRI based research (the activity in the parieto-frontal cortical circuit for grasping is dependent on the context) are put against the background of the findings with single cell recordings. A claim is made about the mirror neuron system, but the explanation is based on
mirror neurons. However, the exact relation between mirror neurons and a mirror neuron system remains implicit. The working of the mirror neuron system is in need of more clarification than the mirroring function of neurons can provide. We have argued that it is unlikely that individual neurons are capable of mirroring highly abstract categories of actions or action goals. So explaining how a mirror neuron system is capable of detecting intentions remains an open challenge.

To be sure, we certainly do not wish to dispute the findings of Iacoboni et al. After all, people can infer intentions from a context, so obviously this is a capacity the brain has and Iacoboni et al. have shown that the motor system is involved in this. However, we do object to the depiction of this capacity as mirroring. Also, this is not to say that, for example, broadly congruent mirror neurons of type 3 are not involved in intention detection, but rather that they are unlikely to provide a similar solid basis as in the case of strictly congruent mirror neurons. Results on action recognition are in need of further explanation in order to support theories on action understanding. Single cell recordings cannot carry by themselves the full weight of support for theories on action understanding.

Conclusion

Debating something trivial as the name of neurons might at a first glance come across as a futile matter, not worth the entire representational analysis. But it is important to note that this is more than just quibbling about the right name for these neurons. The working of the neurons does naturally not depend on the name we use for describing them. However, grouping various kinds of neurons under the label mirror neuron can obscure important differences regarding their functional contributions to the recognition of actions and understanding of goals. When, for example, some mirror neurons are supposed to do more than just mirroring, this has serious consequences for, for example, the idea that the existence of mirror neurons supports the direct matching hypothesis.

Mirror neurons are almost always regarded as carriers of representations, but the wide range of representational claims about mirror neurons gives rise to conceptual difficulties. On the basis of the above analyses, one could argue that there is hardly any problem in claiming that strictly congruent neurons can be said to mirror, as their representational content reflects a property on the relatively concrete and observable level of movements. Also, the representational content of group 1 broadly congruent neurons can be deemed to be on a relatively low and unproblematic level of abstraction. More debatable is the proper interpretation of the activity of group 2 of the broadly congruent neurons. The categories, hand action versus non-hand action are of a rather abstract level, as many different actions can be grouped together within such categories.

The level of abstraction involving goals of actions utilized to describe neurons in group 3 as being mirror neurons seems to be of such a general nature that claims on direct observation without inference become dubious. The retrieving of a goal involves an evaluation of the context, the actor, past experiences with similar situations etc. This is a lot of processing to be described as merely mirroring.

This certainly does not mean that group 3 of the broadly congruent neurons and the non-congruent neurons are less interesting or less important in the coupling of perception to action. On the contrary, as strictly congruent mirror neurons represent the action observed or executed and a single neuron is not capable of recognizing an action by itself, the crucial steps in action recognition must already have been taken. Neurons from the broadly congruent or non-congruent category may very well be part of this action recognition mechanism, representing partial contributions to unfinished results of the analysis.

When analyzing the human mirror neuron system, researchers often use data from single cell studies in their explanation of how the system is able to facilitate action recognition, action understanding or goal recognition. However, when it is problematic to state that individual neurons mirror categories of actions or action goals, as we have argued, single cell study findings cannot play the key role in explaining the human mirror neuron system without further clarification.

Acknowledgments

The authors wish to thank Iris van Rooij for her constructive comments. The present study was supported by a NICI-internal graduation grant to the second and last author, and the EU-Project Joint Action Science and Technology (IST-FP6-003747) and a NWO-VICI grant to the last author.

References


1 As Dinstein et al. (2008) point out, mirror neurons as defined in the single cell studies are very difficult to establish in humans.


