

A Philosophical Reflection on Mirror Neurons: How Mirror Neurons Support Simulation Theory

Nalini Elisa Ramlakhan (naliniramlakhan@email.carleton.ca)

Institute of Cognitive Science, 1125 Colonel By Drive
Ottawa, ON K1S 5B6 Canada

Keywords: simulation theory; theory-theory; mirror neurons; theory of mind; mindreading

Introduction

The discovery of mirror neurons in macaque monkeys has sparked a lot of excitement and debate in neuroscience, cognitive science, psychology, and philosophy. Mirror neurons were discovered in the premotor cortex of macaque monkeys and termed mirror neurons because the neurons “respond both when a particular action is performed by the recorded monkey and when the same action, performed by another individual, is observed” (Gallese & Goldman, 1998). Some psychologists and neuroscientists, such as Marco Iacoboni (2009), claim that mirror neurons help solve the problem of other minds. That is, the problem of how we access and understand the minds of others around us. Similarly, many philosophers also theorize about how we access and understand other minds.

It is generally agreed upon in philosophy and psychology that humans have mind-reading abilities (the ability to explain and predict others’ behaviour). What is not agreed upon is which theory best describes how humans come to have these abilities. On this topic, there is much debate. Some philosophers, such as Simon Baron-Cohen (1995) endorse theory-theory. Here, mindreading is thought of as a Theory of Mind Module (ToMM) in the Fodorian sense (a looser conception is thinking of a module as a neurocognitive mechanism). In this theory, all humans are equipped with a ToMM. In the same way that we all have a language faculty, we all have a ToMM, and it works in much the same way as the language faculty. According to critics of theory-theory, this picture is flawed because we appeal to unobservable laws to explain and predict human behaviour, where mental states that predict others’ behaviour arise from “theoretical reasoning involving tacitly known causal laws” (Gallese & Goldman, 1998). Many developmental psychologists and functionalist philosophers of mind accept theory-theory as an accurate picture of how humans come to have mind-reading abilities. An equally popular and opposing theory to theory-theory is simulation theory. Simulation theorists argue that in order to understand, explain, and predict others’ behaviour, individuals mentally put themselves in the shoes of another person and imagine how they would feel in the given context. The main difference between the two theories is that theory-theory attempts to explain mind-reading abilities as arising from unknown tacit laws, whereas simulation

theory attempts to explain mind-reading as impersonating or mimicking the target. There are other theories that attempt to explain how humans come to have mind-reading abilities, but our main focus here will be on simulation theory, while touching briefly on theory-theory. We will discuss these theories in more detail below.

With the discovery of mirror neurons, some philosophers, such as simulation theorist Alvin Goldman, put forth the hypothesis that mirror neurons support simulation theory over other theories such as theory-theory. He argues that given the role that mirror neurons play in macaque monkeys and possibly humans, mirror neurons accord well with simulation theory, but not other theories (Gallese & Goldman, 1998). More specifically, Gallese and Goldman (1998) argue that “the activity of mirror neurons, and the fact that observers undergo motor facilitation in the same muscular groups as those utilized by target agents, are findings that accord well with simulation theory but would not be predicted by theory-theory.” Given what we know about mirror neurons, I believe that Goldman’s picture of mirror neurons providing support for simulation theory is accurate. In this paper, I argue that there is ample evidence to support the view that the existence of mirror neurons accord well with simulation theory. An in-depth look at the function of mirror neurons and simulation theory shows that mirror neurons work hand in hand with simulation theory and can be viewed as empirical evidence for simulation theory, and supports simulation theory more so than theory-theory. Furthermore, off-line simulation, which will be discussed below, works automatically and like a heuristic, in much the same manner that mirror neurons do, showing that mirror neurons may be an underlying feature of simulating the behaviour of others. Before I defend the view that mirror neurons support simulation theory, it will be useful to begin with a closer look at what exactly mirror neurons are and how they function, and an explication of how simulation theory gives rise to mind-reading abilities.

Before beginning, I would like to note that this paper looks at theories of theory-theory proposed by functionalist philosophers of mind and developmental psychologists, such as Baron-Cohen. There are other versions of theory-theory that differ from these versions. However, due to the space and focus of this paper, I will be looking at only one version of theory-theory, and conclude alongside Goldman that “the activity of mirror neurons, and the fact that observers undergo motor facilitation in the same muscular groups as those utilized by target agents, are findings that accord well with simulation theory, but would not be

predicted by theory-theory” (Gallese & Goldman, 1998). It is also important to note that I am not wholeheartedly rejecting theory-theory as a plausible theory. Rather, the aim of this paper is to show that mirror neurons support Goldman’s theory of simulation. This paper is exploratory in nature, and the main contribution is to emphasize that theories should be supported with ample empirical evidence, such as neural or behavioural evidence, when it is available, and should be examined and analyzed with the findings from cognitive science and related disciplines.

Mirror neurons

Simply put, mirror neurons are neurons that are activated when an individual performs an action and when an individual witnesses someone else performing the same or a similar action. For instance, when I reach for an apple and when I see you reach for an apple, my brain evokes similar neuronal responses. That is, my mirror neurons are activated when I reach for an apple and when I see you reach for an apple (or reach for something else). However, this is not to say that mirror neurons are always active simply when we look at individuals perform an action. For instance, if I observe you randomly waving your arm around without an intention, I will not activate the same neuronal response if I were to wave my arm around. This is because mirror neurons discharge during a goal-oriented action (Gallese and Goldman, 1998). As Gallese and Goldman (1998) point out, “all MN (mirror neurons)...discharge during specific goal-related motor acts. Grasping, manipulating and holding objects are by far the most effective actions triggering their motor response.” A similar view of mirror neurons is taken by Iacoboni, and it is no surprise that the premotor cortex, where mirror neurons are said to reside in the macaque brain, is the region important for “the planning, preparation, and selection of movements and coordinated actions” (Iacoboni, 2009). Furthermore, Iacoboni (2009) points out that within area F5 of the macaque brain (one of the two main fields of the ventral section of the premotor cortex), there are neurons that discharge not only when the monkey performs goal-oriented actions...but also when the monkey, completely still, simply observes somebody else performing these actions.” It is also important to keep in mind that only one third of mirror neurons fire for the *same* executed action and observed action. The other two thirds fire for executed actions and observed actions that achieve similar or related goals, but are not necessarily the same (Iacoboni, 2009). This will be especially important when we look at whether mirror neurons support simulation theory.

Before moving on to simulation theory, it is important to note that one possible function of the mirror neuron system is to promote learning through imitation. However, Goldman and Gallese think that there is more to the mirror neuron system than imitation alone, and that it may be the case that mirror neurons are a rudimentary part of mind-reading abilities.

Simulation theory

According to Goldman (1989), simulation theory is an attractive one. In order to explain and predict human behaviour, individuals “ascribe mental states to others by pretending or imagining themselves to be in the other’s shoes, constructing or generating the (further) state that they would be in, and ascribing that state to the other. In short, we *simulate* the situation of others, and interpret them accordingly” (Goldman, 1989). Goldman also points out that simulation cannot be used in a simplistic manner, where we generalize about how others may feel without taking into account individual differences. For instance, if you attempt to determine my next move in chess, you will probably imagine yourself with my goals and level of chess sophistication. If I am an amateur and you are a master at chess, you will need to not only put yourself in my shoes in order to determine my next move, but also take into consideration how horrible I am at chess. You will then infer my next chess move, but not play it. This is because when simulating, you create the pretend beliefs and desires that you think I have, and these are fed into your decision-making mechanism. Instead of acting on those decisions, the simulation is taken off-line and used to predict my behaviour (Gallese & Goldman, 1998). This process is termed off-line simulation. Furthermore, since we do not always consider such factors or may lack knowledge about certain individuals, simulation can account for why we misinterpret someone’s behaviour. For example, if you walk into a room and see me crying, you may first assume that I am sad or upset when I may really be shedding tears of joy.

It is important to note that theory-theory accounts for these sorts of predictions as well. It should not be assumed that I am taking the position that theory-theory is incompetent in explaining or predicting human behaviour. Rather, I am attempting to show that mirror neurons better support simulation theory than theory-theory. This is mostly due to the empirical nature of simulation theory. Simulation theory is empirical in nature due to the origins of simulation theory, and the evidence used to support it. As will be shown below, mirror neurons accord well with this theory.

Furthermore, Goldman claims that simulation is not the only method used to explain human behaviour. There are other ways to explain and predict behaviour, such as by means of induction. If I am always cheerful when you see me, you can assume I will be cheerful the next time you see me. If I am a shy individual, you can assume that I may smile politely or not say anything at all the next time you see me. The point here is that “simulation is an intensively used heuristic, and one on which interpretation fundamentally rests” (Goldman 1989). If simulation is indeed a heuristic, we can think of simulation as occurring quickly and automatically, and may not be the lengthy process it is initially made out to be, especially if we can predict the behaviour of others through means of induction. We can simulate the behaviour of others, then, rather quickly, although it may appear as if putting yourself in the

shoes of another person to explain or predict their behaviour can be a lengthy process. What is more, many heuristics are not necessarily innate features, but are a combination of capacities and learning. For example, in many connectionist networks, a Hebbian style framework is employed, where upon learning new tasks and repeating them, the neurons that are fired together end up being wired together. This allows for automaticity and reflexivity. Another model that holds a similar view while not necessarily endorsing a Hebbian framework is Iacoboni's ideomotor model framework. These models do not concern us here. However, it is important to note that simulation can be thought of as a heuristic rather than a very lengthy process. For instance, Goldman (1989) asks us to consider the following case initially posed by Kahneman and Tversky:

Mr. Crane and Mr. Tees were scheduled to leave the airport on different flights, at the same time. They traveled from town in the same limousine, were caught in a traffic jam, and arrived at the airport 30 minutes after the scheduled departure time of their flights. Mr. Crane is told that his flight left on time. Mr. Tees is told that his was delayed, and just left five minutes ago. Who is more upset?

We usually answer the question instantaneously, empathizing more with Mr. Tees, knowing that he would be more upset than Mr. Crane. We reach the conclusion by simulating how we would feel if we were Mr. Tees or Mr. Crane, and respond according to those feelings. Again, we don't spend a whole lot of time mentally putting ourselves in their shoes. Rather, the process of simulating is quick and effortless. Another example is refraining from telling certain jokes if we know someone may find it offensive, such as telling a religious joke when the person sitting across from you is wearing a large cross around their neck. We could infer that if we were in that person's shoes, we wouldn't find the joke funny and may possibly find it rude.

A final and related note about simulation theory: simulation occurs quite naturally, sometimes when we are not even aware of it. For example, if you're watching a boxing or tennis match, you will probably simulate the behaviour of whoever you are rooting for. It is a natural occurrence, as Goldman (1989) points out, because "simulation need not be an introspectively vivid affair. The approach can certainly insist that most simulation is semi-automatic, with relatively little salient phenomenology."

Simulation theory vs. theory-theory

As other theories of mind-reading abilities do not concern us here, I will not discuss them. However, I will briefly mention the difference between simulation theory and theory-theory to provide further clarification of simulation theory. Recall the examples given above about predicting chess moves or predicting which passenger would be more upset. While simulation theory claims that we mimic or attempt to understand how that person would feel or what they would do in the given context, theory-theory "does not utilize any pretend states that mimic those

of the target" (Gallese & Goldman, 1998), and the attributor does not use her decision-making mechanism to arrive at a prediction of the other's behaviour, thoughts, or feelings. This is important because as Gallese and Goldman (1998) point out, "if there is evidence for mental mimicry in the mind-reading process, that would comport nicely with ST (simulation theory) and would not be predicted by TT (theory-theory)." Theory-theory predicts that the apparatus of intentional explanation consists in the possession of laws. However, Goldman (1989) claims that this theory only shows that individuals can grasp mental laws, but does not show that these laws are innate or mentalistic. We will now turn our attention to determining whether mirror neurons support simulation theory.

How mirror neurons support simulation theory

Given what we know about mirror neurons and simulation theory, there is enough evidence to suggest that mirror neurons accord well with simulation theory. Iacoboni and Goldman claim that understanding other minds is an effortless activity and involves some sort of imitation. Recall that mirror neurons are firing patterns that occur both when an individual performs an action and when an individual watches someone else perform the same action or a similar action. The neurons are, in a sense, mimicking the actions of the person being observed while not necessarily causing an output, that is, not causing any motor activity in the observer to perform the action. This seems to accord well with off-line simulation described above. When we simulate the behaviour of others and predict their behaviour, we don't have the same behavioural output ourselves. For instance, if I observe you stretching your arm and notice that your hand is in a grasping motion, I can predict that you will be reaching for something. Given the description of mirror neurons above, mirror neurons fire when we witness an action that we have previously performed that achieves the same or a similar goal. If these neurons are fired when we observe goal-oriented actions that are the same or similar to the ones we have previously performed without executing the action, we can interpret this as support for simulation theory. Recall that Goldman claims that in off-line simulation, we don't execute the action when we simulate, but rather the simulation is taken off-line and we instead predict the behaviour of the other person. If mirror neurons do in fact create a mental state in an observer that resembles or matches that of the person being observed without executing the action, we can view this as a neural event that underlies the event of simulating. As Gallese and Goldman (1998) make the point: "externally generated MN (mirror neuron) activity does not normally produce motor execution of the plan in question. Externally generated plans are largely inhibited, or taken 'off-line', precisely as ST postulates." We can view mirror neurons as the neural basis for simulation theory, where mirror neurons can be thought of as "nature's way of getting the observer into the same 'mental shoes' as the target—exactly what the conjectured

simulation heuristic aims to do” (Gallese & Goldman, 1998). Thus, it would seem that we have some sort of empirical support for simulation theory.

Because mirror neurons fire when we observe a person perform a goal-oriented action, we can predict what the goal is. When we simulate, the same process occurs. More specifically, we can predict the goal of the observed person by placing ourselves in their shoes and attempt to determine what we would do in the given context. If we attempt to explain the behaviour of another person in a given context when we have yet to be placed in such a context or do not have similar experiences that the observed person has, it may be rather difficult for us to simulate. This is why many of us have a hard time predicting the behaviour of serial killers or schizophrenics. We can only put ourselves in their shoes to a certain extent, as we do not have much knowledge of their subjective experience, although we can attempt to learn what it is like to be one of those individuals.

Mirror neurons appear to provide more support for simulation theory than the competing theory, theory-theory. Gallese and Goldman (1998) claim that “MN activity is not mere theoretical inference. It creates in the observer a state that matches that of the target. This is how it resembles the simulation heuristic. Nothing about TT leads us to expect this kind of matching.” Recall the description of theory-theory given above. Theory-theory claims that the mental states that we attribute to others are unobservable and are theoretical posits that arise from tacitly known casual laws. Furthermore, theory-theory does not postulate that mimicry has much to say about how we come to explain and predict others’ behaviour. The point here is not to show that theory-theory is an inadequate theory. Rather, the point here is to provide further support for the idea that mirror neurons accord well with simulation theory. Whereas theory-theory does not appear to have an empirical basis and does not attempt to look at the neural underpinnings of mind-reading, simulation theory does. Simulation theory is supported empirically by mirror neurons, corroborating what is already known about mind-reading abilities. However, this does not imply that mirror neurons cannot or do not support theory-theory. Although theory-theory posits that mind-reading abilities emerge from tacitly known causal laws and are unobservable, it is obviously not the case that they dismiss any neural underpinnings that would give rise to such abilities. Rather, theory-theory theorists appear to be more concerned with theoretical posits and unobservable laws. Furthermore, both theory-theory and simulation theory are concerned with cognitive empathy. That is, they are concerned with understanding how individuals have the capacity of knowing what another individual believes, feels, or intends by placing themselves in another person’s situation (Kauppinen, forthcoming). And as Kaplan and Iacoboni (2006) point out, there is a correlation between cognitive empathy and mirror neuron activity. This could mean that either theory-theory or simulation theory is correct and supported by mirror neurons. Insofar as simulation theory emphasizes the importance of mimicry

and imitation and appears to be more empirically oriented than theory-theory, we can reasonably conclude that mirror neurons do in fact support simulation theory, more so than theory-theory. What is more, as Goldman and Gallese (1998) point out, mirror neuron activity is “no mere theoretical inference. It creates in the observer a state that matches that of the target. This is how it resembles the simulation heuristic. Nothing about TT leads up to expect this kind of matching.”

Before concluding, I would like to make a brief point about mirror neurons and mimicry. It may be the case that mirror neurons do not actually give rise to any sort of mind-reading abilities. Rather, it may be the case that mirror neurons are needed to mimic or to imitate but cannot give rise to representing another individual’s mental state (see Decety, 2011). Even if this is the case, which I do not think that it is, mirror neurons can be viewed as a precursor to mind-reading abilities, in the same way it can be viewed as a precursor to affective empathy. This does not suggest that mirror neurons no longer support simulation theory. Mimicry and imitation are important factors in developing mind-reading abilities. Furthermore, even if mirror neurons were merely a precursor to predicting the mental states of others, it still provides an empirical basis for simulation theory. We first need to experience similar actions and be placed in similar contexts in order to predict another person’s goal directed behaviour. When mirror neurons fire during the observation of another individual performing an action that we have previously performed, mirror neurons are mimicking the behaviour without actually performing it. According to Iacoboni, this is what allows us to understand the actions of another individual—we are able to infer the goal of the observed individual because we have previously performed the same or a similar action. Even if mirror neurons do not go so far in explaining the behaviour of another individual, the ability to mimic the behaviour without performing it still seems necessary to eventually simulate the other person’s behaviour. However, this does not suggest that the mirror neuron system should be viewed solely as a function for imitative abilities.

Conclusion

Determining how individuals acquire mind-reading abilities initially began as a theoretical journey. Reflecting from the arm chair gave rise to various theories about how individuals predict and explain others’ behaviour in an attempt to dissolve the problem of other minds. With reflections from various mind-reading studies in psychology, philosophers were better able to position their theories. The discovery of mirror neurons sparked more discussion and debate amongst philosophers of mind and psychologists studying cognitive empathy, where some scholars argued that mirror neurons do not give rise to mind-reading abilities while others argued that it supports their theory. Alvin Goldman falls into the second camp. Throughout the course of this paper, I have attempted to

defend Goldman's view that mirror neurons support simulation theory. I have shown that mirror neurons accord well with simulation theory and there are empirical reasons to believe that they work hand in hand with the ability to predict and explain others' behaviour. Furthermore, I have argued that while it is reasonable to suppose that mirror neurons support simulation theory, we should not disregard the possibility that mirror neurons may also support other theories of mind-reading, such as theory-theory. However, I have attempted to show that the discovery and function of mirror neurons rests better with simulation theory.

In order to successfully conclude whether mirror neurons support simulation theory, further research must be completed. While this paper has focused on a specific version of theory-theory, it should not be assumed that all versions of theory-theory take for granted what has been described in my analysis of theory-theory. There are other theories that have accounts of goal-oriented behaviour that appear to be more empirical in nature (see Maibom, 2003). Due to focus and space constraints, these theories have not been looked at in this paper. However, it will be important to examine and analyze these theories in order to conclude whether mirror neurons support simulation theory more so than theory-theory. For instance, Gopnik (1993) looks to developmental evidence as support for theory-theory, rather than incorporating both psychological and neuroscientific evidence. An examination of other versions of theory-theory, as well as the developmental evidence needs further analysis before concluding that mirror neurons support simulation theory. It may turn out that mirror neurons support the developmental evidence for theory-theory, in which case, we would need to be hesitant in making a hasty conclusion.

References

- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press
- Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review*, 3, 92-108.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2, 493-501.
- Goldman, A. (1989). Interpretation psychologized. *Mind and Language* 4, 161-185.
- Gopnik, A. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behavioural and Brain Sciences*, 16, 1-14.
- Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653-670.
- Kaplan, J.T., & Iacoboni, M. (2006). Getting a grip on other minds: mirror neurons, intention understanding, and cognitive empathy. *Social Neuroscience* 1, 175-183.
- Kauppinen, A. (forthcoming). Empathy and the moral sentiments.

Maibom, H. (2003). The mindreader and the scientist. *Mind and Language* 18, 296-315.