Mirroring, mindreading and smart behaviour-reading

Article

Accepted Version

Borg, E. ORCID: https://orcid.org/0000-0003-2725-9568 (2017) Mirroring, mindreading and smart behaviour-reading. Journal of Consciousness Studies, 24 (5-6). pp. 24-49. ISSN 1355-8250 Available at https://centaur.reading.ac.uk/67081/

Published version at: http://www.ingentaconnect.com/contentone/imp/jcs/2017/00000024/F0020005/art00002

Publisher: Imprint Academic

Accepted for publication in *Journal of Consciousness Studies*.
Anticipated publication 2017.

Mirroring, mindreading and smart behaviour-reading

Emma Borg
Philosophy Department
University of Reading, UK
e.g.n.borg@reading.ac.uk

Abstract

This paper examines the claim that mirror neuron activity is the mechanism by which we come to know about the action-related intentions of others (e.g. Gallese et al 1996, Rizzolatti et al 2009), i.e. that they are a mechanism for 'mindreading'. I agree with recent authors (e.g. Hickok 2008, Jacob 2008) who reject this view but nevertheless I argue that mirror neurons may still have a role to play in the ways in which we understand one another (social cognition). If we adopt a certain kind of pluralism about social cognition then the mirror neuron system could play a role in social cognition *even if* it provides no access to the minds of others at all. I argue for this view and consider what the approach might entail for the ontology of the mirror neuron system.

Accepted for publication in *Journal of Consciousness Studies*.
Anticipated publication 2017.

Mirroring, mindreading and smart behaviour-reading[*]

Mirror neurons (MNs) are neurons which fire in two distinct conditions:

- the performance by a subject, S, of a given motor act (e.g. a precision grip of a cup).
- the witnessing by S of a target conspecific performing that action.

Since mirror neurons were first discovered (Gallese et al 1996) they have generated a great deal of interest and substantial claims have been made about the role that they might play in cognition. Specifically (as we will see in §1) they have been widely promoted as providing the neural basis of action understanding – as being the mechanism by which we come to understand that someone else is grasping a glass with the intention of drinking or with the intention of moving the glass elsewhere, for example. However there has also been a growing backlash in philosophy of mind and cognitive science against this view of the mirror neuron system. Claims such as that made by Ramachandran 2000, 2011 that MNs provided 'the driving force behind the "great leap forward" in human evolution' and that their discovery will 'do for psychology what the discovery of DNA did for biology' now, I think it is fair to say, seem over-hyped. However, even if this pessimism about mirror neurons as the seat of intention understanding is right (as I think it is), a further question remains: if mirror neurons don't underpin intention-recognition, just what do they do for us? The discovery that our brains are, in certain respects, doing the same thing whether we pick up a cup or we witness a conspecific pick up a cup, is a surprising and fascinating finding. So, if we reject the claim that MNs play a role in mindreading, is there anything we can say about the possible role of the MN system in humans?[1] In this paper I want to suggest that there is, that there can still be a role for MNs to play within social cognition even if turns out they are not the mechanism by which we come to know the minds of others.[2]

To get this result, I'll suggest that we need to adopt the kind of pluralist view of social cognition proposed by Andrews 2012 and Feibich and Coltheart 2015 (to be explored in §2). I'll

---

[*] Thanks are due to colleagues in the Reading Centre for Cognition Research for helpful discussion. Research for this paper was supported by a John Templeton Foundation award for work on 'Mirror neurons, intersubjectivity and intellectual humility'.

[1] Some authors have expressed doubts about the very existence of a human (as opposed to a monkey) mirror neuron system; see, e.g., Chong et al 2008, Dinstein 2008, Lingnau et al 2009. However, whilst accepting that some degree of caution may be proper here given the different forms of evidence primarily available in each domain – with invasive procedures yielding single cell recordings possible for monkeys but, for the most part, only less direct evidence, such as fMRI, available for humans (though see Mukamel et al 2010 for an exception) – I will set this most sceptical stance to one side and simply assume in what follows that the existence of the human mirror system is well-established. The question for this paper can thus be seen as doubly conditional: *if* there is a human mirror neuron system and *if* it doesn't underpin intention-recognition, *then* what does it do? I'll touch on the reasons for thinking that the antecedent of the second conditional holds below (§1) but for proper discussion, see Borg 2007, 2013, Hickok 2008, Jacob 2008.

[2] This paper thus seeks to answer the question posed in Borg 2007: 'If MNs are the answer, what was the question?' I'll suggest that the sort of question that MNs might be the answer to is 'what are the mechanisms by which we acquire, store and utilise certain learnt behavioural associations?'

argue that any pluralist view should draw a fundamental distinction between acts of social cognition which depend on mental state attribution and acts of social cognition which rely on what I'll term 'smart behaviour-reading'. The mistake of much of the mirror neuron literature, I'll contend, has been to hold that mirror neurons contribute to some form of mental state access or attribution, whereas a more conservative proposal is that the system serves to track behavioural regularities.[3] I'll suggest (§3) that this positive view of the role mirror neurons play in social cognition deserves serious consideration. Finally, this will lead us to ask about the ontology of the MN system and I'll rehearse evidence for what seems to be the standard claim that the system should be viewed as largely innate (§4). However, I'll suggest (§5) that this evidence for innateness is questionable and that, on the contrary, we have good reason to think (as the statistical behaviour-reading view suggests) that the MN system is largely (perhaps entirely) composed of acquired capacities. Finally, I will conclude by suggesting that this finding is illustrative: if we adopt pluralism about the mechanisms of social cognition and maintain that non-mentalistic tracking of behaviour is important, this part of the social cognition system will be best understood as acquired. On this model, much of our understanding of the actions of others will be seen to rely on a subject's history, building on their unique background of experience and interactions. For behaviour-reading, I'll suggest, we predict what we have lived.

1) Mirror neurons as a mechanism for intention recognition

The surprising finding – that our brains are (in certain respects) in the same state whether we are ϕ-ing or are witnessing someone else ϕ-ing – has led many theorists to suggest that mirror neurons provide the neural mechanism underpinning our understanding of the intentional states of others. For instance, as Gallese and Goldman 1998: 497 initially proposed:

> [I]t is conceivable that externally-generated MN activity serves the purpose of 'retrodicting' the target's mental state, moving backwards from the observed action. Let us interpret internally generated activation in MNs as constituting a plan to execute a certain action, for example, the action of holding a certain object, grasping it or manipulating it. When the same MNs are externally activated – by observing a target agent execute the same action – MN activation still constitutes a plan to execute this action. But in the latter case the subject of MN activity knows (visually) that the observed target is concurrently performing this very action. So we assume that he 'tags' the plan in question as belonging to the target.

Following this model, a wide range of researchers have claimed that the MN system underpins our recognition of the intentions behind the actions performed by others. For instance, Rizzolatti, Fogassi and Gallese 2009: 636 claim 'there is little doubt that the mirror mechanism is one of the most basic and possibly the most basic mechanism for intention understanding',

---

[3] This view of mirror neurons has been defended at length in work by Heyes, e.g. Heyes 2005.

Iacoboni et al 2005 argue that MNs should be understood not merely as capturing the *what* of an action but its *why* (i.e. the intention with which a given motor act was performed), Buccino et al 2007: T119 state that 'the MNS [mirror neuron system] appears [also] to underlie the coding of intentions behind the actions of others' and Ocampo and Kritikos 2011: 262 claim that MN activity allows 'an observer to distinguish between an agent's ultimate intentions'. [4] A common view amongst those working on mirror neurons thus seems to be:

The MN account of understanding of intentional action:

> We come to understand the intention behind the action of another through mirroring/motor resonance; e.g. your grip of the glass triggers a motor plan for drinking in my MN system, I 'tag' this plan as attaching to you. So seeing your movements allows me to know directly about your (action-related) mental states.

Although, as noted, many theorists have found this kind of account attractive, there has been, I think it is fair to say, something of a growing backlash against it. One apparent problem concerns the underdetermination of intentions by actions: if MNs are held to respond to overt kinematic gestures (such as the precision grip of the cup) their activity will underdetermine intention attribution, for a target may perform one and the same action with a variety of different intentions. For instance, I might grip a cup to drink, or to place, or to throw (see Borg 2007, Jacob 2008). This worry has sometimes been labelled the 'intentional worry' for the MN approach (for discussion see, e.g., Borg 2013). In response advocates of the MN approach have argued that this simple objection does not in fact hold. For although some MNs (labelled 'strictly congruent') do fire only with respect to specific kinematic gestures, others ('broadly congruent' ones) are held to show a sensitivity to the overall action goal which the specific kinematic gesture subserves. So, for instance, Fogassi et al 2005 apparently show that undertaking extremely similar kinematic gestures (whereby a monkey reaches for a piece of food, grasps it and either moves it to their mouth, to eat, or moves it to their shoulder, to place in a container there) led to different patterns of MN firing as early as the reaching stage. While Rochat et al 2010 (building on a paradigm developed by Umiltà et al 2008) showed that very different sets of kinematic gestures (e.g. operating pliers versus 'reverse pliers') led to the same pattern of MN firing when the different kinematic gestures subserved the same overall action goal (e.g. object grasping).[5] The

---

[4] Finally, concentrating on the idea that MNs underpin simulation (which in turn underpins social cognition) Gallese 2000 writes: 'MN activity seems to be nature's way of getting the observer into the same 'mental shoes' as the target, which is exactly what the conjectured simulation heuristic is all about', see also Blakemore & Decety 2001.

[5] See also Kohler et al 2002 where the sound of nuts being cracked was sufficient for triggering MNs associated with the motor act of cracking nuts.

suggestion has thus been that while patterns of firing for a single MN may not capture action goals, MNs cluster to form chains of motor organisation and it is the activation of such a chain that forms recognition (or enaction in the subject's own case) of an action goal. As Sinigaglia 2008: 84 writes:

> [I]t is due to this motor chain organisation that grasping is not just grasping for grasping's sake, but is grasping to carry food to the mouth and eat, or grasping-to-move X from A to B, etc….It is…a question of recognising the role of chains of motor goals, and the fact that they shape the motor expertise, which is at the base both of the agent's capacity to act and to represent his/her actions and his/her ability to understand the immediate significance of the actions and intentions of others.

Whether or not the appeal to MN chains is able to solve the intentional worry remains contentious (for representative negative views, see Hickock 2008, Borg 2013) but, rather than rehearsing those extant arguments here, I want instead to suggest that we take a step back and re-examine what is involved in the fundamental premise that MNs have a role to play in supporting our interaction with, and understanding of, one another (i.e. in supporting social cognition). For, if we accept a certain kind of pluralism about the systems underpinning social cognition, it seems that we might allow that MNs have a role to play in social cognition (i.e. in the way in which we come to predict and explain the actions of others) *without* admitting that they give us access to the minds of others at all (thus simply avoiding objections such as the intentional worry altogether, as it would turn out that MNs were *not* involved in giving us access to the mental states of others after all). It is this proposal that I want to explore in the rest of the paper.

2) Pluralism about social cognition

In the 1990's and early 2000's in philosophy of mind and cognitive science, the fundamental debate about our access to other minds concerned the plausibility or otherwise of theory-theory accounts versus simulation theory models. According to the former, typically developed adults are in possession of a theory of mind (something they either acquired, e.g. Gopnik and Meltzoff 1997, or have innately, e.g. Scholl and Leslie 1999), which consists of a set of folk psychological rules (perhaps grouped inside a mental module) along the lines of 'if A wants p and believes that doing q will bring about p then, ceteris paribus, A will do q'. To predict or explain the action of another, then, the theory-theory view states that the subject assigns a set of mental states to the target (e.g. wanting p and believing 'if q then p') and uses the rules of their folk psychological theory to work out what the target will do next or why she did what she did. On the opposing view, advocates of simulation-based approaches (e.g. Heal 1996, Gordon 1996, Goldman 2006) suggest that subjects use their *own* internal planning mechanisms in an imaginative exercise. To explain or predict your behaviour I 'put myself in your shoes' and think what I would do if I were in your position and had the kind of mental states that I take you to have. The philosophical debate between theory-theory versus simulation-based accounts did not get decisively settled one way or the other, though a kind of consensus grew up that a complete account of social cognition would probably involve both approaches (see Goldman 2006).

This move towards some kind of hybrid account concerning the mechanisms of social cognition opens the door for a more wide ranging kind of pluralism, anticipating the idea that social cognition might rely on a potentially wide range of different mechanisms. This idea has now been pursued at length by Andrews 2012 and Feibich & Coltheart 2015, who argue that there are (at least) two different ways in which subjects might come to understand the actions of others:

i) <u>Mental state attribution</u>: A witnesses B's behavioural and environmental cues and uses these to ascertain B's mental state. A uses these attributed mental states to predict and explain B's behaviour. E.g.:
    - Jill is wearing running clothes and appears breathless and red in the face. She is filling a glass from the tap. A infers: Jill is thirsty and wants a drink. A predicts: Jill will drink from the glass she is filling.

ii) <u>Smart behaviour-reading</u>: A witnesses B's behavioural and environmental cues and predicts and explains B's behaviour on the basis of these cues. E.g.
    - Jill is wearing running clothes and appears breathless and red in the face. She is filling a glass from the tap. A knows: the most likely next action in presence of these cues is that the target drinks from the glass. A predicts: Jill will drink from the glass she is filling.

(i) captures a kind of standard folk psychological approach to action prediction and explanation, maintaining that we come to understand another's actions by seeing the agent through the lens of mental state attribution; we 'mentalise' the subject and take the posited mental states to underpin the action in question. (i) is quiet on whether we engage in this mental state attribution via an explicit bit of reasoning utilising a subject's theory of mind or whether we come to consider the mental states of others via the kind of imaginative exercise posited by (at least explicit versions of) simulation theory, but however we do it, (i) claims that (at least some) social interaction involves viewing a target as in possession of mental states. It is far from clear, however, whether or not (ii) also makes this commitment. To see this, let's examine smart behaviour-reading in a little more detail.

The idea that behaviour-reading might matter for social cognition has emerged recently within two distinct research strands: on the one hand, within the philosophy of mind, there has been growing interest in this kind of account amongst theorists who favour an 'embodied', 'embedded', 'enactive' or 'extended' approach to cognition (for discussion of so-called '4e' approaches, see, e.g., Rowlands 2010). On this bundle of approaches, thinking is held not to involve (or at least, not to be exhausted by) the computational manipulation of abstract symbols in brain systems. Rather thinking is held to involve the whole situated agent, drawing in (in some fundamental sense apparently not accommodated by standard Cartesian models) the agent's bodily states (beyond the brain) and involving her wider environment. Within these kinds of non-Cartesian approaches, then, social cognition is viewed as involving a sensitivity to, and perhaps resonance with, the motor behaviour and environmental cues of others.

The second area where behaviour-reading accounts have come to the fore is in the domain of non-human animal research. For in response to Premack and Woodruff's 1978 question 'Do chimpanzees have a theory of mind?', Daniel Povinelli and colleagues have argued at length that any non-human behaviour which at first blush might seem to reveal reasoning about unobservable mental states, in fact can and should be explained by treating the animal in question as a smart behaviour-reader instead. So, for instance, advocates of smart behaviour-reading suggest that the fact that chimpanzees beg only from others who can see them should be explained, not by taking the chimpanzee to be reasoning about what another can or can't see (mental state), but reasoning about observable, non-mental features of the situation such as whether it is possible to draw an uninterrupted line from the open eyes of the other to the self. In the animal literature, then, the idea that (ii) provides an *alternative* to mental state ascription models is clear. However, in the philosophical literature it is less clear that this is the case. Take, for instance, a McDowell-style view (McDowell 1996), where we are encouraged to reject the schism Descartes (apparently) introduced between mind and world, operating instead with a properly 're-enchanted' model of the natural world. Assuming a behaviour-reading approach within this kind of account may not obviate the appeal to mental state attribution. Rather the claim of (ii) within this sort of model will be that sensitivity to behaviour gives us direct, non-inferential access to the mental states of others. Following the later Wittgenstein, the idea here will be that we simply *see* the intentional states of others in their embodied actions.

I want to suggest then that there are two distinct ways in which an appeal to smart behaviour-reading can be understood:

> ii.a) Behaviour-reading provides direct, unmediated access to the inner lives of others. It is not the case that we witness the behaviour and environment of another agent and use

this as evidence in an inferential act which delivers an explicit attribution of mental states to the target. Rather sensitivity to the embodied actions of others gives us perceptual access to their mental states, which we may represent in entirely motor terms.

ii.b) Behaviour-reading provides a way to predict and explain action without *any* access to or representation of the mental states or inner lives of others.

On (ii.a) the method of social cognition is still mentalising: we come to understand, predict and explain the actions of others through knowing about the mental states that led to the production of given behaviour. The difference between (i) and (ii) on this reading then lies, first, in the route to knowledge of those mental states: in (i) we infer mental states from behavioural cues, in (ii) no inference is needed, we perceive them directly through sensitivity to the situated actions of others. Secondly, (ii.a) makes no claims about how knowledge of the inner lives of others are represented by a target. The beliefs and desires of others *may* be represented explicitly (as when I believe that Bill is thirsty or that Deepak wants the glass to be filled) but they may also be represented more implicitly, as in embodied views where my knowledge of the mental states of others comes about through formation of a motor plan which realises a given mental state in my own case and the subsequent tagging of that motor plan as belonging to another.

In (ii.b) however the model is different: on (ii.b) we predict and explain the actions of others simply via appeal to behaviour with no claim that this gives us any kind of access to the inner lives of others (however these inner states may be represented). On (ii.b) action prediction is all 'surface level' as it were. Given this interpretation, the difference between (i) – mental state attribution – and (ii) – smart behaviour-reading – is that (i) requires viewing others through the lens of unobservable mental states (however these mental states are represented by the subject), while (ii) requires only an awareness of typical patterns of behaviour, no appeal to mental states, intentions or 'inner lives' is involved.[6]

Looking back at the claims made on behalf of MNs in underpinning social cognition, it seems that the MN account of understanding intentional action belongs to (ii.a): on the proposed view, we get at the intentional states of others not through complex inferential work and explicit reasoning about unobservable mental states. Rather we know about the intentional states of others directly through motor resonance with their actions, where this kind of understanding of other's mental states need not involve explicit representation of those mental states as such (i.e.

---

[6] For further discussion of the difference between mental state attribution and smart behaviour-reading, see AUTHOR (manuscript).

thinking a propositional thought such as *Bill wants a drink* is not necessary in order to understand that Bill wants to drink from the glass he is holding).[7] For instance, Gallese 2007: 659 writes:

> …social cognition is not only 'social metacognition'; that is, *explicitly* thinking about the contents of someone else's mind by means of symbols or other representations in propositional format. We can certainly 'explain' the behaviour of others by using our complex and sophisticated mentalising abilities. And we should add that the neural mechanism underpinning such complex mentalising abilities are far from being fully understood. Most of the time, though, we do not need this. We have a much more direct access to the inner world of others. Direct understanding does not require explanation. This particular dimension of social cognition is embodied, in that it mediates between the multimodal experiential knowledge of our own lived body and the way we experience others.[8]

The view I want to recommend in this paper, on the other hand, is (ii.b). This is not a claim about how we access or represent the inner lives of others, rather it is the claim that *no* access to or consideration of the inner lives of others, in any form, is required for (a wide range of) action understanding. On (ii.b) the claim is that we often make predictions about the way someone else will behave based entirely on statistical regularities about behaviour. These are mechanistic predictions of the kind that could, in principle, be made for things which lack any inner life at all. We predict that someone will raise a glass to their mouth to drink after grasping that glass simply because that is the statistically most likely next action in the given context – something we have learnt from our experience of observing people interact with objects like these in situations like this. Although this prediction may utilise a specialist, dedicated neural mechanism (the MN system) for recording, storing and responding to the behavioural regularities of others, in principle the very same kind of prediction could be made by us for robotic arms, coke machines, and falling objects, things which we don't feel any compulsion to attribute intentional states to at all.[9] So, how plausible is (ii.b) as a model of what MNs are doing for us?

---

[7] So, is it really right that standard MN accounts sign up to (ii.a), rather than (ii.b)? (I'm grateful to a referee for raising this question.) It does seem to me that the quotes supplied at the start of section 1 make it look as though it is (ii.a) rather than (ii.b) which is on the table, as does the alignment of MN accounts with wider embodied theories of cognition more generally (for instance, Caramazza et al 2014: 3 refer to the mirror neuron theory as "perhaps the most influential embodied theory of cognition"). However, perhaps reading MN accounts in this way is a mistake and advocates should in fact be read as signing up to (ii.b) after all. If this is right what such accounts must mean by their talk of 'direct access to the inner lives of others' is access to the internal but purely causal motor systems of others, with no claim that this introduces any genuinely *intentional* element into the picture at all. Furthermore, the notion of 'action goal' in play must be one which applies equally to human and animal action, and to movements by inanimate objects such as coke machines, cash dispensers and falling rocks (for a good discussion of the ambiguity surrounding the term 'action goal', see Hickok 2008, Goldman 2009: 239). This isn't how I've read standard MN accounts, but of course if this is the right reading of them then what follows should be seen as an attempt to clarify this aspect of extant accounts rather than being read as an objection to them.

[8] I'm grateful to a reviewer for highlighting this passage.

[9] So why might there be a specialised neural system for recording and responding to the behavioural regularities of other humans, if in principle what the system is sensitive to are features shared by other non-human objects? A possible answer to this question may lie in the fundamental (evolutionary) salience that specifically human action has for us, together with the fact that, due to our shared motor systems, it proves possible to co-opt the motor system in this way.

3) Mirroring as a mechanism for acquiring, storing and applying behavioural regularities

The model I want to suggest for MNs then is that they provide a mechanism not for accessing the mental states of others but for recording and recovering typical human behaviour in certain kinds of environmental contexts. What chains of MN activation might be capable of doing for us is allowing direct, non-mentalised access to the likely movement a target will perform next, because chains of motor neurons could capture a kind of brute expectation about the way kinematic gestures usually hang together.[10] Like a system for tracking word frequency which knows which pairs or strings of words are most likely to follow, given the start of a sentence, but which make no appeal to the semantics of words, so MN activation might yield a prediction about the next probable action even though they make no prediction about the intentions of the target performing the act.[11] Furthermore, it seems clear that a system underpinning easy associative learning of behavioural regularities would be very useful in a range of social cognition situations. For in cases of simple co-ordination (e.g. avoiding people in corridors or passing another person an object), or in contexts which are highly stereotyped (e.g. driving, playing sport, purchasing items in shops, ordering from menus, etc.), it seems highly unlikely that we need to access the intentional states of our social partners. When predicting whether another driver will go round a cyclist or wait behind them, or working out whether a tennis partner will serve down the line or out wide, it seems plausible to think that I am responding to subtle, probably sub-consciously registered, kinematic clues rather than engaging in deductive reasoning about the other person's beliefs and desires. In this way it seems we might rely on purely non-mentalising mechanisms, simply associating behavioural cues with outcomes.[12] Thus I want to agree with Lurz et al 2014: 446 when they write:

---

[10] As a reviewer pointed out, this way of phrasing the point is controversial, because (as noted above with respect to work by Rochat and others), the claim is that MNs respond to action outcomes *not* kinematic gestures. Two points on this however: first, the relationship between MNs and kinematic gestures remains controversial at least according to some authors (e.g. Borg 2007, 2013, Hickok 2008). Second, it is possible that action outcomes could be understood entirely in terms of 'the way kinematic gestures hang together'. For instance, an action outcome of grasping an object, say, or drinking, could be rendered in terms of the final kinematic gesture (e.g. the closing of fingers round the object or the movement of the open lips to the glass) in relation to earlier gestures (the extension of the arm, etc.). Of course, this would require a many:one mapping from sets of gestures to final gesture (different precise movements could lead up to the final fingers-close-on-object, etc) but construed in this way the system would still be generating an expectation about the way that kinematic gestures usually hang together, where this expectation involved no intentional understanding of those kinematic acts.

[11] As 2013 notes, this understanding of MN activity is entirely consistent with what Hickok 2008: 1231 calls 'the null hypothesis': 'namely that F5 is fundamentally a motor area that is capable of supporting sensory-motor associations'.

[12] The idea that in such predictions the system is akin to that used for automated systems is also found in Andrews 2012: 51: 'I predict the person's behaviour using the same reasoning as I use for the machine, by making generalizations from past experience and knowing the target's function'. Though note that Andrews also argues for a greater demotion of belief/desire reasoning than is suggested in this paper (e.g. 2012: Ch.6).

> It is quite obvious that when predicting an opponent's behaviour in a sporting match, humans use a quick and effortless behaviour-reading mechanism, and that in highly stereotypical social interactions (e.g. ordering meals at restaurants, or buying food at a butcher shop) humans employ behavioural rules that range over representations of social roles and deontic rules.

We have a picture then of one of the mechanisms for social interaction as resting not on some complex, inferential mental state attribution and an application of general (perhaps innate) folk psychological rules, nor on a non-inferential, embodied understanding of another's inner life, but rather as a response to the bodily behaviour of another which merely exploits past regularities of behaviour to which we have been exposed.[13] In this way, the MN system would be useful in action prediction, and would thus have a role to play in social cognition, even though it is not in any way a mechanism for understanding the minds of other people or their inner, intentional states. If this were to prove a feasible way to think about what MNs do for us, then it would turn out that advocates of MNs were right to posit a role for MNs in supporting social cognition, though they were wrong to think that this entailed that MN activity is a mechanism for mindreading. However adopting this view of the role of mirror neurons in social cognition would, it seems, have repercussions for other claims made about the mirror neuron system, in particular concerning the ontology of the system.

4) Mirror neuron chains as innate

It is, I think, fair to say that the received view is that the mirror neuron system is largely an innate endowment. For instance, see work by Rizzolatti and Arbib 1998, Rizzolatti and Craighero 2004, Fogassi et al 2005, Iacoboni et al 2005, Ferrari et al 2006, Johnson 2005, Lepage and Théoret 2007, Park et al 2009, Pascalis and Kelly 2009, all of whom appear to endorse an innateness hypothesis. According to this picture, there are prewired motor chains embedded in the MN system and the combination of witnessing a target perform a small-scale action (e.g. precision grip of a cup) together with the context in which the action occurs (e.g. in a 'before tea' versus an 'after tea' setting, see Iacoboni et al 2005) results in the triggering of one or other of these prewired chains. The claim that a significant part of the MN system is innate might be thought warranted by considering one of the main pieces of evidence for the existence of a MN system in humans, namely spontaneous neonatal mirroring (by an infant of a caregiver). So, in the classic

---

[13] One question now might concern the difference between this kind of action prediction/explanation and the prediction/explanation we engage in for non-intentional objects – what is the difference between predicting that my opponent will serve the tennis ball out wide and predicting that the boulder will roll down the cliff? I think an appeal to MNs might give us an interesting answer here, for establishing the boundaries of the things which activate the MN system would give us one way to draw the boundary between things we are willing to treat as minded versus non-minded. Though MN activation does not directly involve intentional understanding, it might nevertheless underpin our pretheoretical categorisation of the world into intentional and non-intentional objects.

experiment from Meltzoff and Moore 1997, infants were found to be capable of mirroring the facial gestures of others (such as tongue protrusion) within weeks of birth, and these findings have been supported more recently by work apparently showing mirroring of other caregiver gestures, such as finger protrusion (see Nagy et al 2005) and also work showing neonatal imitation amongst monkeys (e.g. Ferrari et al 2006). The extremely early-onset of this mirroring behaviour has been taken to support the innate endowment view of the MN system.[14] Thus Simpson et al 2014 argue:

> While experience, including associative processes, undoubtedly plays a role in developing the corpus of behaviors that infants imitate…an associative learning account of the fundamental capacity to imitate is incompatible with the evidence on two fronts. First, only minutes to hours after birth, human infants imitate opening and closing of eyes…, head movements…, the /a/ sound…, index finger protrusion…, facial gestures (e.g., mouth opening, tongue protrusion); …, and emotional facial expressions (e.g., happiness, sadness, surprise…) prior to having opportunities to form strong associative links between action observation and imitative responses. Similarly, macaque infants reared in a nursery from birth imitate before they have experienced any contingent facial interactions with caregivers…, and they additionally show specific electroencephalogram changes (i.e., mu suppression), evidence of a functioning MNS, on the day of birth... These results fail to support an associative learning account of neonatal imitation.

Furthermore, neonatal mirroring of facial gestures provides a conceptual argument – sometimes labelled 'the Correspondence Problem' (e.g. Brass and Heyes 2005) – for innateness. The Correspondence Problem starts by noting that mirroring appears to play an important developmental role. For instance, Ferrari et al 2009 argue that facial imitation by macaques in the first few days of life are predictive of the infant monkey's motor skill development during the first year of life. Mirroring, then, it is argued, has a special role to play in longitudinal social development.[15] The problem is that, in order to play a developmental role, an infant must be able to tell when they are being mirrored, yet it is unclear how such awareness could arise since the infant lacks visual feedback for its own facial gestures (unlike, say, its own arm gestures). What we must explain, then, is how the infant is capable of recognising that a caregiver smile mirrors a self-produced smile, when only one side of the equation – the seen-smile of the caregiver – is visually presented. One clear way this explanatory need could be met, and the Correspondence Problem solved, would be if the infant came equipped with an innate system – the MN system – whose job was to pair, from birth, seen and felt gestures.[16] In this way, an infant would not need

---

[14] Casile et al 2011 also point to the distinct developmental trajectory for facial mirror neurons vs. hand mirror neurons to support the argument that the former, unlike the latter, might be present from birth.

[15] See Paukner et al (forthcoming).

[16] As Casile et al 2011: 8 write: '[T]he main question is, how does the infant translate the observed movement into a correspondent motor program without a direct visual experience of the own face? One possible explanation might be that both face processing and the mirror neuron system, or at least the part involved in facial movements, rely on a brain network that is present already at birth and whose basic elements are probably genetically predetermined (Ferrari and others 2006; Johnson 2005; Lepage and Théoret 2007; Park and others 2009; Pascalis and Kelly 2009). In other words, part of the visual information related to facial movements appears to be matched already at birth with the corresponding motor representation'.

to learn that the caregiver smile mirrored its own smile since the operation of an innate mirror system would underpin knowledge of this matching.

So, should we accept the view that (at least) some significant part of the MN system is innate and that MN chains are prewired parts of our cognitive endowment? I want to suggest not. A first point to note is that the evidence in favour of the innateness hypothesis is open to question. For instance, the appeal to neo-natal mirroring highlights one of the problems with discussion of the mirror neuron system in general, for 'mirroring' is used in the literature to describe very different kinds of activity (and thus as supported by different kinds of evidence). So mirroring is used to talk about:

i. Mirror neuron activity.
ii. Activation of matching muscles (EMG, MEP).
iii. Observable-mimicry (matching of bodily movements, perhaps vocal imitation).

The best evidence for (i) is single cell recording, however as an invasive procedure this is not standardly available for humans (though it may be possible in cases where cognitive impairment licenses invasive brain procedures on other grounds). Somewhat less conclusive evidence for (i) is available from brain scanning procedures, such as fMRI and MRI. Evidence for (ii) comes from measurements of muscle potentialities, for instance EMG recording of Zygomaticus Major activity in response to witnessing a smiling face (see Sims et al 2012). Finally, evidence for (iii) is available from simple, surface level observation of the agent. It is assumed, then, that these three kinds of behaviour are all instances of the operation of a single underlying system – the MN system – and thus that evidence of (ii) or (iii) can be taken as evidence of (i). However, this is something we could question, especially as the degree of cognitive control, or the penetration of higher-level content, seems very different in each case. For instance, it has been shown that both behavioural mirroring (iii) and muscle mirroring (ii) are sensitive to highly abstract features of the perceived situation such as the relative social status of participants, whether the target is associated with rewarding or unrewarding behaviour, and a subject's overall aims or desires.[17] While activation of MNs (i) is presumably a much more automatic response, triggered primarily by observable, e.g. motor, aspects of the viewed scene. Presumably, if my brain does the same thing when I undertake a precision grip and when I witness such a grip, this neural mirroring is independent of concerns such as the social status of the observed target.[18]

---

[17] So, for instance, we mirror the bodily posture of those we perceive as being of a higher social status than ourselves, or those we wish to construct social alliances with (see, e.g., Winkielman et al 2011) and typically developed adults show greater muscle activation for mirroring of facial gestures with respect to high-reward over low-reward faces (see Sims et al 2012).

[18] I'm grateful to a reviewer for pointing out that this is actually a controversial point, since some studies do indeed suggest that MN activity itself is just as sensitive to abstract features of a situation as are behavioural mirroring and

A second point to note is that claims of neonatal mirroring by-infant-of-adult remain highly contentious, with concerns that the original findings from Meltzoff and Moore have proved hard to replicate and meta-analyses of the data apparently undermining claims of mirroring behaviour (see Jones 2009 and the references therein; for response see Simpson et al 2014, followed by Heyes 2015 for counter-response. Finally Oostenbroek et al 2016, which reports the largest longitudinal study of neonatal imitation in humans to date, finds no evidence of neonatal mirroring and suggests that earlier apparent evidence of mimicry might be an artefact of restricted sample sizes). Furthermore, as many theorists have pointed out, the kinds of gestures in question here (particularly tongue protrusion) are ones performed more frequently by infants in states of arousal, such as when they are receiving dedicated adult interaction. Thus the experimental situation itself is one which will increase the prevalence of these gestures, raising the possibility of random matching of behaviour. It seems then that the best we can say is that the jury remains out on whether an appeal to observable neonatal 'mimicry' can support the positing of an innate MN system (for a more strongly negative conclusion, see Heyes 2016).

It is also unclear that the Correspondence Problem really forces us to endorse an innate model of MN mechanisms either. For instance, very careful coding of facial mirroring (by-infant-of-adult) situations would be needed to show that there is absolutely no difference in either the type or the intensity of feedback that the infant receives when mirroring gestures occur (as opposed to other types of social interaction), as any such difference would be able to explain infant awareness of mirroring without recourse to an innate mechanism. Alternatively, one might think that a kind of bootstrapping is possible whereby, during the earliest stages of development, the infant is engaged in creating an internal body image map, pairing seen and felt bodily gestures to one another (e.g. pairing the feeling of moving its own arm thus-and-so with the sight of either its own or another's arm moving thus-and-so). It is at least conceivable that building up this seen and felt body map, with its implicit representation of relational facts (e.g. the spatial relation between fingers, hands and arms) could eventuate in an inferential jump from 'felt bodily

---

muscle mirroring (e.g. Sobhani et al 2012, which argues for the influence of political allegiance on MN activation). On the one hand, however, if turns out that MN activity is modulated by complex, abstract features of a viewed scene this would, I take it, provide support for the main claim of this section, namely, that MN responses relating to action understanding are acquired rather than innate (since I take it positing an innate system which included representation of very abstract features like political ideas would be unattractive). On the other hand, if it turns out that MN activity is indeed modulated by a wide range of complex contextual features (like kin relationship, political ideas, social standing, etc) then I wonder if this might not also risk undermining the model of MN function advocated by Rizzolatti, Gallese, and others, for it would require MN activation to occur downstream of a complex, abstract representation of a situation. Construed in this way, although the MN system could play a role in understanding others, it would no longer be as a system which provided some kind of direct, motor-generated access to the mental states of others, for the activation of the system would now depend on how the target was viewed by the subject in terms of things like political affiliation. Motor competencies alone would not be in the driving seat for understanding others, rather these motor responses would be modulated by our higher-order beliefs.

movement here (own face)' to seen bodily movement at an appropriate spatial location (face of other). In this way even without an innate system for matching seen and felt movements the infant would be able to come to recognise that, for instance, a seen smile matched a felt smile.

So I would suggest that neither the direct appeal to neo-natal mirroring nor the Correspondence Problem suffice to show that an innate model of the MN system is required. In addition, however, we should also be clear that regardless of how the controversy around neonatal imitation is finally settled, arguments in support of an innate system for facial mirroring don't carry over in any direct way to claims about an innate system for mirroring of other kinds intentional actions. Crudely, even were it might ultimately to turn out that there is an innate system for recognising and responding (via imitation) to faces, this wouldn't entail that the system for motor imitation of non-facial bodily gestures is similarly innate. Yet it is non-facial gestures which the MN account of understanding of intentional action relies on and thus far we have no reason to think that this must be similarly innate. Finally, as discussed in the next section, it also seems that we have positive reasons to think that at least this part of the MN system reflects learnt associations.

5) Non-facial mirror neuron chains as acquired

On an alternative view of the MN system, rather than MN responses forming part of our innate cognitive endowment they are held to be acquired through learning (see Cook et al 2014, Heyes 2015). The suggestion is that a subject witnesses specific actions which lead to a given behavioural outcome during a learning phase. Habituation to these relations (in context) then leads to anticipation of the habituated outcome when the subject is later exposed to the same kind of behavioural stimuli in contexts which sufficiently resemble those encountered during the learning period. So why might we think MN activity is acquired rather than innate? One potential piece of evidence in favour of this model comes from the unpleasant finding that if monkeys are deprived of the sight of their own hands from birth they are later unable to perform visually guided hand movements (Held and Bauer 1967). This suggests that mapping between one's own seen and felt movements is something acquired through early exposure (though it is worth noting that Held and Bauer's finding is also consistent with the idea of an innate mechanism which requires early environmental triggering).

Stronger evidence in favour of the acquisition view, however, comes from experimental evidence for substantial MN plasticity. For instance, Heyes et al 2005 repeatedly exposed subjects to novel action pairs, e.g. subjects were instructed to raise their little finger in response to the sight of a middle finger being raised. After training (repeated instances of raising one's own little finger in response to the sight of a raising middle finger), it was found that the sight of a middle finger raise was enough to trigger the MN response associated with the subject's raising of her little finger. This acquisition of an incongruous MN response through training shows that MN activity patterns are sensitive to environmental factors and can be altered through learning. This kind of MN plasticity provides good evidence, I think, in favour of viewing mirror neuron activity as something we acquire through exposure to behavioural contingencies.[19] Furthermore, I think this finding is illustrative: if we are interested in the role behaviour-reading plays in social cognition it seems extremely likely that the knowledge underpinning this aspect of thought will be acquired rather than innate. For the vast majority of actions that we need to predict and explain are mediated by the kind of variable social mores which only a learning account could accommodate. Whether one's group has a practice of regularly eating food as soon as it is grasped or has a practice of waiting until all relevant conspecifics have food in hand before starting to eat one's own food will, on a behaviour-reading account, determine what prediction you now make when faced with a conspecific grasping food. Associative learning of regular behavioural contingencies will reflect the kind of experiences the agent has had in the past and the kind of social group to which she belongs, rather than some innate universally shared inheritance.[20]

Finally, it seems that this account fits well with the proposed view about the function of the MN system within social cognition. For if the role of MNs is to provide us with an efficient way to record and track behavioural regularities (by invoking motor matching for actions rather than subsuming the behaviour under any kind of mental state perspective), it seems to make sense that this motor information would need to be learnt rather than innate. The chains we would form would be the result of our lived experience rather than something we would come equipped with from birth. Furthermore, this seems to fit with what we know of the role of learning and familiarity in the MN system and with general findings about neural plasticity. As is now well-known, the MN system responds differently given the degree of motor-familiarity one

---

[19] This stance also fits with evidence about MN activity in response to tool use, as Casile et al 2011: 3 note: 'It was originally reported that F5 mirror neurons do not respond if the same action (e.g., grasping) is performed with a tool. However, more recent results showed that after extensive visual exposure of the monkey to actions executed with a tool, a subset of mirror neurons in the ventral part of area F5c started responding also to this type of visual stimuli (Ferrari and others 2005)'.

[20] As noted, this point would sit entirely comfortably with a view of the ontogeny of the MN system whereby it was only the MN system for facial imitation which was innate, with other parts learnt. I'm grateful to a reviewer for stressing this point and highlighting comments, e.g. by Gallese et al 2009: 106, which suggest that this is the view advocates of MN accounts intend. It is however perhaps also worth noting that if this is the kind of nativist view proposed by advocates of standard MN accounts, then it is unclear why associationist learning views such as those advocated by Heyes et al have traditionally been seen as *opposed* to standard views of the MN system.

has with an action (i.e. whether or not one has performed the action oneself, and if so, how often). For instance, in a range of influential works, it was shown both that the MN systems of expert dancers respond more strongly to witnessed dance moves than do the MN systems of non-dancers and that the MN systems of trained dancers respond selectively to the form of dancing in which the subject is trained. That is to say, the MN systems of professional dancers respond more strongly to witnessed dance moves than do the MN systems of the general populace, while amongst professional dancers those trained in capoeira, say, respond more strongly to capoeira moves than they do to ballet moves (see, e.g., Calvo-Merino et al 2005). What all this points to, I think, is an understanding of mirror neurons as a mechanism for acquiring and applying learnt behavioural contingencies, rather than as an innate system capable of providing a window into the mind of another. Following from discussion in §2 though, this does not entail that MN activity has no role to play in the kind of thinking that deserves the label of 'social cognition', i.e. the kind of thinking we do when we are trying to predict or explain the actions of others. For, on the learnt behaviour-reading model, we can still maintain a role for MNs in social cognition along the lines of (ii.b) (though not, as advocates of mirror neurons often see to suggest, as (ii.a)). In understanding the actions of others I suggest we often focus exclusively on the *action* neglecting entirely the *other*. We often have no need to engage in mental state attribution or consideration because, in a range of situations (e.g. scripted exchanges in shops, playing sports, avoiding others in the street, regular patterns of behaviour amongst those we know) we make do simply with statistical regularities about learnt behavioural contingencies. I can predict you are going to drink from the glass you are filling simply because this is the kind of action that normally follows that kind of behaviour in this kind of context – I don't need to consider your inner life in any way, shape or form. If this model of the role of MNs is right, it would be a mistake to think that the MN system plays any part in the philosopher's notion of mindreading or intention attribution, but equally it would be a mistake to think that all social engagement involved mindreading or intentional understanding. MNs, as part of our motor competence, don't support our understanding of other minds though they may still support our prediction of other's action.

6) Conclusion

It is common to claim both that mirror neuron activity underpins our access to the action intentions of others (i.e. that MNs have a special role to play in our access to the intentional, mental states to others) and that, because of the special role that action imitation plays in our lives, the mirror neuron system responsible for matching seen and felt movement must be (at least in large part) innate. This paper rejects both of these claims. As many authors have now argued, there are serious problems to be faced by the view that mirror neuron activity underpins intention recognition, however, as I have sought to show here (§2) this need not entail that mirror neuron activity has no role to play within social cognition. For our thinking about others often involves nothing more than associationist prediction of ensuing actions. On seeing a target ϕ we predict that they will δ because δ-ing is what we have come to associate with ϕ-ing in this kind of context.[21] This kind of thinking oils the wheels of social interaction, allowing us to engage easily with one another in a wide of range of contexts, but it invokes no special mentalising skills, it requires no mental state attribution or mysterious embodied access to the 'other'. Mirror neuron activity, I want to suggest, might reflect just these kinds of learnt behavioural contingencies and might in this way contribute to social cognition, even whilst we reject claims about the place of mirror neurons in mindreading. One repercussion of this way of thinking about what mirror neurons do for us is that it makes an innateness model of the mirror neuron system (at least for non-facial gestures) unattractive (since we are responding to learnt contingencies not innate matching mechanisms), but I have tried to suggest that this in itself is an attractive feature of the model for we have good reason to think that mirror neuron activity is indeed the result of experience rather than the product of an innate inter-modal matching system.

Bibliography

AUTHOR, manuscript. On deflationary accounts of social cognition.
Andrews, K. 2012. *Do Apes Read Minds?* Cambridge, MA: MIT Press.
Blakemore, S. & J. Decety. 2001. From the perception of action to the understanding of intention. *Nature Reviews Neuroscience* 2: 561-567.
Borg, E. 2007. If mirror neurons are the answer, what was the question? *Journal of Consciousness Studies* 14: 5-19
Borg, E. 2013. More questions for mirror neurons. *Consciousness and Cognition* 22: 1122-1131
Brass, M. & C. Heyes. 2005. Imitation: is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Science* 9: 489-495.

---

[21] This interpretation of the role of mirror neurons would also be consistent with the kind of predictive coding view of the mind advocated by Clark (e.g. Clark 2013), where neural systems are constantly anticipating expected input.

Calvo-Merino, B., Glaser, D. Grèzes, J., Passingham, R. and Haggard, P. 2005. Action observation and acquired motor skills: an FMRI study with expert dancers**.** *Cereb Cortex*, 15: 1243-1249.

Caramazza, A., Anzellotti, A., Strand, L., & Lingau, A. 2014. Embodied cognition and mirror neurons: a critical assessment. *Annual Review of Neuroscience* 37: 1-15.

Casile, A. Caggiano, V. and P. F. Ferrari. 2011. The mirror neuron system: a fresh view. *Neuroscientist:* 1-15.

Catmur, C., Walsh, V., & Heyes, C. 2007. Sensorimotor learning configures the human mirror system. *Current Biology* 17: 1527-1531.

Clark, A. 2013. Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behavioural and Brain Sciences* 36: 181-204.

Craighero, L., 2014. The role of the motor system in cognitive functions. In L. Shapiro (ed.), *The Routledge Handbook of Embodied Cognition*, London: Routledge.

Cook, R., Bird, G., Catmur, C., Press, C. & Heyes, C. 2014. Mirror neurons: from origin to function. *Behavioural and Brain Sciences,* 37, 177-241.

Dinstein, I. 2008. Human cortex: reflections of mirror neurons. *Current Biology* 18: R956-R959

Ferrari P., Visalberghi E., Paukner A., Fogassi L., Ruggiero A., et al. 2006. Neonatal imitation in rhesus macaques. PLoS Biol 4: e302.

Ferrari P., Paukner A., Ruggiero A., Darcey L., Unbehagen S., Suomi S. 2009. Interindividual differences in neonatal imitation and the development of action chains in rhesus macaques. *Child Dev* 80:1057–68.

Fiebich, A. & Coltheart, M. 2015. Various ways to understand other minds: towards a pluralistic approach to the explanation of social understanding. *Mind and Language* 30: 235-258.

Fogassi, L., Ferrari, P, Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. 2005. Parietal lobe: from action organisation to intention understanding. *Science* 308: 662-667.

Gallagher, S. 2005. *How the body shapes the mind*. Oxford: OUP.

Gallagher, S. 2009. Neural simulation and social cognition. In J. A. Pineda (Ed.), *Mirror neuron systems. N*ew York: Humana, 1-17.

Gallese, V. 2001. The 'Shared Manifold' hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies* 8: 33-50.

Gallese, V. 2007. Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. Philosophical Transactions of the Royal Society B 362: 659–669.

Gallese, V. and Goldman, A. 1998. Mirror neurons and the simulation theory of mind reading, *Trends in Cognitive Science* 2: 493-501.

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. 1996. Action recognition in the premotor cortex. *Brain* 119: 593-609.

Goldman, A. 2006. *Simulating Minds*. Oxford: Oxford University Press.

Goldman, A. 2009. Mirroring, simulating and mindreading. *Mind and Language* 24: 235-252.

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

Gordon, R. 1996. Radical simulation. In: Carruthers, P. and Smith, P. (eds.) *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Heyes, C. 2005. Imitation by association. In S. Hurley, N. Chater (eds) *Perspectives on Imitation: from mirror neurons to memes*. MIT Press.

Heyes, C. 2010. Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews* 34: 575-583

Heyes, C. 2015. Homo imitans? Seven reasons why imitation couldn't possibly be associative. *Philosophical Transactions of the Royal Society: B. 371: 20150069*

Heyes, C. 2016. Imitation: not in our genes. *Current Biology* 26: 412-414.Heal, J. 1986. Replication and functionalism. In *Language, Mind, and Logic*, J. Butterfield (ed.), Cambridge: Cambridge University Press.

Held R, and Bauer J. 1967. Visually guided reaching in infant monkeys after restricted rearing. *Science* 155: 718–20

Hickok, G. 2008. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience* 21: 1229-1243.

Iacoboni, M., I. Molnar-Szakacs, V. Gallese, G. Buccino, J. Mazziotta, G. Rizzolatti. 2005. Grasping the Intentions of others with one's own mirror neuron system. *PLoS Biology* 3, e79: 0001-0007.

Jacob, P. 2008. What do mirror neurons contribute to human social cognition? *Mind and Language* 23: 190-223.

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 USA* 106: 9925–9930

McDowell, J. 1996. *Mind and World*. Oxford: OUP.

Menary, R. 2008. Embodied Narratives. *Journal of Consciousness Studies*, 15: 63-84.

Mukamel, R., Ekstrom, A., Kaplan, J., Iacoboni, M., Fried, I. 2010. Single-neuron responses in humans during execution and observation of actions. *Current Biology* 20: 750-756.Nagy, E. et al. 2005. Index Finger Movement Imitation by Human Neonates: Motivation, Learning, and Left-Hand Preference. *Pediatric Research* **58**: 749–753.

Ocampo, B. & Kritikos, A. 2011. Interpreting actions: the goal behind mirror neuron function. *Brain Research Reviews* 67: 260-267.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, J., Clark, S., Slaughter, V. 2016. Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology* 26: 1334-1338.

Paukner, A., Simpson, E., Ferrari, P., Mrozek, T. & Suomi, S. J. forthcoming. Neonatal imitation predicts how infants engage with faces. *Developmental Science*

Premack, D. and Woodruff, G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1: 515–526.

Ramachandran, V.S. 2000. Mirror Neurons and imitation learning as the driving force behind 'the great leap forward' in human evolution. *Edge:* 69.

Ramachandran, V.S. 2011. *The Tell-Tale Brain*. Heinemann Press.

Rizzolatti, G. & Craighero, L. 2004. The mirror-neuron system. *Annual Review of Neuroscience* 27: 169-192.

Rowlands, M. 2010. *The New Science of the Mind.* MIT Press.

Scholl, B. and Leslie, A. 1999. Modularity, development and 'Theory of Mind'. *Mind and Language* 14: 131–53.

Simpson, E., Murray, L., Paukner, A. and Ferrari, P.F. 2014. The mirror neuron system as revealed through neonatal imitation: presence from birth, predictive power and evidence of plasticity. *Phil. Trans. R. Soc. B* 369: 1471-2970.

Sobhani M., Fox G., Kaplan J., Aziz-Zadeh, L. 2012. Interpersonal liking modulates motor-related neural regions. *PLoS ONE* 7: e46809.

Sims, T., Van Reekum, C., Johnstone, T. and Chakrabarti, B. 2012. How reward modulates mimicry: EMG evidence of greater automatic facial mimicry of more rewarding faces. *Psychophysiology* 49: 998-1004

Sinigaglia, C. 2008. Mirror neurons: this is the question. *Journal of Consciousness Studies*, 15: 70-92.

Umiltà, M., Escola, L., Intskirveli, I., Grammont, F., Rochat, M., Caruana, F., Jezzini, A., Gallese, V., & Rizzolatti, G. 2008. How pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences* 105: 2209-2213.

Winkielman, P., Berridge, K., & Sher, S. 2011. Emotion, consciousness, and social behavior. In J. Decety & J. T. Cacioppo (Eds.), *Handbook of Social Neuroscience.* Oxford University Press. 195-211.