Unbroken mirrors: challenging a theory of Autism

Victoria Southgate¹ and Antonia F. de C. Hamilton²

¹Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, London, WC1E 7HX, UK
²School of Psychology, University of Nottingham, Nottingham, NG7 2RD, UK

The ‘broken mirror’ theory of autism has received considerable attention far beyond the scientific community. This theory proposes that the varied social-cognitive difficulties characteristic of autism could be explained by dysfunction of the mirror neuron system, thought to play a role in imitation. We examine this theory and argue that explaining typical imitation behavior, and the failure to imitate in autism, requires much more than the mirror neuron system. Furthermore, evidence for the role of the mirror neuron system in autism is weak. We suggest the broken mirror theory of autism is premature and that better cognitive models of social behavior within and beyond the mirror neuron system are required to understand the causes of poor social interaction in autism.

Introduction

Individuals with autism have great difficulty with many aspects of social interaction, but the cause of this disability remains unknown. In recent years the discovery of a mirror neuron system (MNS) in the human brain, made up of regions that respond to the actions of self and other (see Figure 1), has led to an increased interest in the brain systems that underlie basic social processes. The primary function of the MNS is proposed to be related to action interpretation [1], but it also has been implicated in other social-cognitive processes, including imitation [2], theory of mind [3], language [4] and empathy [5]. The broad range of social-cognitive functions attributed to the MNS overlaps to some extent with the various social-cognitive difficulties seen in autism. Thus, there is an intuitive appeal in linking these phenomena and in developing a unified neurocognitive theory of autism. Such a ‘broken mirror’ model has been suggested in several guises in recent years [6–8]. The present paper will examine the data and theories supporting the broken mirror hypothesis of autism and finds several reasons to be cautious of it.

The broken mirror theory

The broken mirror theory of autism has its origins in studies of imitation behavior [6]. Several influential studies suggest a role for mirror neuron regions in hand-action imitation [2,9], and there is also evidence for some degree of imitation deficit in children with autism [10]. On this basis it is argued that dysfunction of the MNS could be the cause of such impaired imitation [6,8]. This model, thus, implies a three-way relationship between a brain system (the MNS), a behavior (imitation) and a disorder (autism) (Figure 2). Some versions of the broken mirror theory take a broader scope than imitation [7] and, drawing on speculations about MNS contributions to empathy [11], theory of mind [3] and language [4], propose that damage to the MNS could cause problems in all these areas [7]. However, in this opinion paper we focus primarily on the evidence for imitation as a key link between the MNS and autism because it is the only social-cognitive ability for which there is both evidence of MNS involvement [2] and some documented deficits in autism [10].

We will first examine the theory and evidence linking mirroring regions of the brain to imitation (Figure 2, arrow a), and imitation to autism (Figure 2, arrow b). We then consider whether these and other studies support a link between mirror neurons and autism (Figure 2, arrow c) and the idea that damage to the MNS should cause specific deficits in social abilities.

Imitation and the MNS

The discovery of neurons in monkeys that respond to both performed and observed actions [1] has led to the identification of a MNS in the human brain (Figure 1), which appears to play a role in imitation. Observation of actions for imitation elicits greater activation in MNS regions than observation without instruction to imitate [2,9,12]. Furthermore, both temporary [13] and permanent lesions [14] to MNS areas cause difficulty with imitation of some actions. Nonetheless, despite the implied involvement of the MNS in imitation, there are at least two reasons for being cautious of the proposal that impaired imitation in autism must stem from a dysfunctional MNS. First, successful imitation goes beyond simply matching actions across bodies, and second, a dysfunctional MNS should be expected to manifest in problems other than imitation.

Imitation is more than mirroring

Some recent models of imitation describe a process limited to the direct transformation of visual information into motor output [2,9,15]. In postulating a dysfunctional MNS as the cause of impaired imitation, broken mirror theorists implicitly endorse this direct mapping. However, successful imitation is not only the matching of corresponding actions from one body to another but also requires several different cognitive processes. These include visual analysis, representation of action goals, selection of what and when to imitate and motor control (Figure 3). Different
imitation tasks might require many types of action representation [16] and varying degrees of action selection [17,18] (Box 1). For example, as described in Box 1, young children evaluate the rationality of an action in selecting whether or not it should be imitated, but this evaluation does not involve the MNS [19].

Furthermore, current data do not support proposals about how the MNS might enable imitation learning [20]. One proposed type of imitation learning involves substituting a new and more efficient means for an old motor act [1]. However, young children sometimes choose to replace an efficient action with an inefficient action [17] or imitate actions that they know to be causally unrelated to the desired outcome [18,21]. Such behaviors are difficult to explain by a process of direct matching but can be accounted for if one considers other factors outside the MNS that are likely to influence imitation (Figure 3). If imitation relies on systems beyond the MNS, it is not sensible to assume a straightforward mapping between imitation behavior and the MNS.

**The MNS is not for imitation alone**

Even those involved in the discovery of mirror neurons and the human MNS agree that the basic function served by the MNS is unlikely to be imitation [1]. For example, macaques have mirror neurons but rarely imitate one another [22], and the human MNS is more active for complementary actions than for imitative actions [23]. Plausible theories suggest the MNS serves the basic function of facilitating action prediction [24,25] or the understanding of action goals [15]. Dysfunction of the MNS should, therefore, impair these abilities, as well as cause poor imitation. In line with this, neuropsychological patients with lesions to MNS regions show poor action interpretation as well as poor imitation [14]. However, the available evidence suggests that individuals with autism do not have difficulties with understanding actions.
[20,26–28]. For example, like typically developing children, children with autism are able to infer the intended goal of a failed action [28]. If the MNS serves the basic function of action interpretation, results such as these suggest that it is not dysfunctional in autism.

**Imitation in autism**

A second component of the broken mirror hypothesis is the claim that children with autism have a specific deficit in imitating actions. Indeed, a recent meta-analysis concludes that imitation difficulties are a core feature of autism [10]. However, this view has recently been challenged [29]. More importantly, if there is no single cognitive process or brain system underlying imitation behavior, then interpreting the reasons for poor imitation performance in children with autism is not straightforward.

**Imitation success and failure in autism**

Children with autism often fail imitation tasks [10], in particular, those requiring imitation of meaningless actions or facial expressions, and for which no explicit instructions to imitate are given. This failure is key to the broken mirror hypothesis, but there are several recent findings suggesting intact imitation in autism, which are not easily accommodated. For example, individuals with autism show an enhanced automatic imitation effect [30] and normal interference effects when observing an incompatible action [31]. Furthermore, children with autism can perform a variety of imitation tasks correctly when they are explicitly instructed to imitate [29,32]. These results are not compatible with the broken mirror proposal that children with autism have a fundamental difficulty in matching the actions of self and other.

The fact that children with autism can imitate but tend not to do so without instruction suggests that their difficulties might arise from problems with knowing when and what to imitate [20,26]. Knowing when and what to imitate probably depends on the ability to exploit the social and communicative cues of others [33]. Typically developing children use the presence of social cues to modulate their imitation [18], but there is some evidence that children with autism do not (Box 1). A reduced sensitivity to social cues, including a lack of preference for looking at the eyes, and infant-directed speech [34,35] is well-documented in autism and could quite plausibly lead to atypical imitation behavior [20] (Figure 3).

**The MNS and autism**

The evidence cited above demonstrates that imitation does not depend solely on the MNS and that the proposed MNS contribution to imitation (matching the actions of self and other) is unlikely to be damaged in autism. Consequently, we suggest that studies of imitation behavior and its relationship to the MNS do not support the broken mirror theory of autism (Figure 2, arrows a and b). However, some data have emerged recently suggesting that neural responses in the MNS of individuals with autism differ from those of control participants. Although these experiments do provide a more direct test of the broken mirror theory (Figure 2, arrow c), unfortunately the available data do not paint a clear picture of MNS activity in autism, being either difficult to interpret or contradictory.

Some of these studies have used indirect measures of neuronal activation in the MNS, including muscle activation [36], excitability of motor cortex [37] or suppression of resting-state rhythms over motor cortex [38]. It is, however, difficult to pinpoint the origin of the atypical activation that these authors attribute to the MNS. For example, reduced resting-state (mu) suppression over sensorimotor cortex in individuals with autism compared with controls could be due to differences in MNS processing [38] but could equally well be due to differences in earlier visual processing. Reduced attention to social stimuli [34], reduced processing of biological motion [39] and differences in more general understanding of complex visual information [40] have all been documented in autism. Because visual systems, in particular those processing biological motion, are a necessary input to the MNS, it is quite plausible that abnormal visual processing in autism could cause abnormal responses within the MNS. Therefore, atypical MNS activity in response to viewing biological...
actions should not be automatically attributed to a dysfunctional MNS.

Several studies have also produced contradictory results. One study reports that individuals with autism lack modulation of primary motor–cortical excitability for viewing hand actions from a ‘self’ perspective (but not ‘other’ perspective) [37]. However, the conclusion that this might reflect a dysfunctional MNS is incompatible with reports that mirroring phenomena are apparent only for viewing actions attributed to another person [41]. Reported fMRI results are also inconsistent, with one study failing to find activation in the inferior frontal gyrus even in control participants [42], contrary to the study on which it was modeled [2].

Studies that have measured both imitation performance and brain activation in MNS regions should provide the clearest evidence for the broken mirror theory [32,42,43]. However, closer examination suggests these results actually constitute evidence against a specific relationship between MNS activation and imitation in autism. Specifically, these studies report atypical activation of MNS areas in individuals with autism during imitation tasks despite the fact that the imitation behavior of these subjects did not differ from control participants [32,42,43]. This strongly suggests that the atypical activation of MNS areas is not related to imitation ability.

To summarize, the evidence for a direct, causal relationship between the MNS regions of the human brain and the social difficulties seen in autism is, at best, weak. Others have also proposed alternative ways of interpreting this data without appealing to a dysfunctional MNS [44]. We also note that even if a perfect experiment revealed a clear relationship between the activation of MNS regions of the autistic brain and imitation behavior, this would not prove that the MNS activity causes poor social cognition in autism, but would leave open the possibility that a lack of attention to social or communicative cues (originating elsewhere in the brain) could cause abnormal responses in the MNS.

**Beyond imitation**

The evidence reviewed above describes why studies of imitation do not make a convincing case for a core MNS dysfunction in autism. Nevertheless, the conceptual link between a brain system for basic social information processing and the social problems seen in autism might remain appealing. Several theories now hint at a broader ‘principle of mirroring’, sometimes expressed as a ‘shared manifold’ or the ‘like-me’ hypothesis [11,45]. The neural substrates of this principle of mirroring would presumably include the traditional mirror neuron regions in frontal and parietal cortex but could extend to any brain regions showing overlapping activation for self and other, including those involved in somatosensory or emotional processes (see Figure 1).

It has been argued that the ability to match any representations relating to self and other, instantiated in the extended mirror neuron regions, is absent in autism and is the primary cause of deficits in social cognition [7]. As yet there is no experimental evidence in support of these broader claims, but there are several reasons to be cautious of such an ambitious but underspecified theory. In particular the critique above highlights the danger of attempting to link brain regions directly to behaviors, and the breakdown of behaviors, without considering the different cognitive processes involved. As we have highlighted, the core MNS regions of the brain are likely to support several cognitive processes, and a breakdown in imitation behavior could arise from the failure of several different cognitive systems either within or outside the MNS (see Figure 3 for a possible model). This same principle continues to apply when an extended MNS or a process of ‘self–other’ mapping is considered. We suggest that the development of a principled and testable model of autism needs to be grounded in an understanding of cognition and how it breaks down [46].

**Concluding remarks**

We have reviewed evidence for links between imitation behavior, the MNS and autism, and find all three sources of potential support for the broken mirror theory are lacking. In particular, it is not yet clear which cognitive components of imitation are supported by a MNS and at which level (e.g. execution, selection) the imitative problems of individuals with autism originate. Other, as yet unanswered questions, have also been raised (Box 2). As such, it is premature to speculate on the involvement of the MNS in autism and, we believe, unwise to promote such a theory in the popular press [47] or to make claims concerning intervention techniques on the basis of such a theory [48].

**Acknowledgements**

A.F. de C.H. is supported by the Economic and Social Research Council grant 061–25–0318 and by a pilot grant from Autism Speaks. V.S. is supported by a European Commission Pathfinder Cooperative Research Grant (CALACE). We thank Coralie Chevallier, Gergely Csibra, Uta Frith and Dana Samson for helpful comments on the manuscript.
References

ARTICLE IN PRESS