Inhibition of imitative behaviour and social cognition

Marcel Brass[1,*], Perrine Ruby[2] and Stephanie Spengler[3]

[1] *Department of Experimental Psychology and Ghent Institute for Functional and Metabolic Imaging, Ghent University, Henri-Dunantlaan 2, 9000 Ghent, Belgium*
[2]*INSERM U821 and Lyon 1 University, 69675 Bron Cedex, Lyon, France*
[3]*Independent Junior Research Group Body and Self, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany*

There is converging evidence that the observation of an action activates a corresponding motor representation in the observer through a 'mirror-matching' mechanism. However, research on such 'shared representations' of perception and action has widely neglected the question of how we can distinguish our own motor intentions from externally triggered motor representations. By investigating the inhibition of imitative response tendencies, as an index for the control of shared representations, we can show that self–other distinction plays a fundamental role in the control of shared representations. Furthermore, we demonstrate that overlapping brain activations can be found in the anterior fronto-median cortex (aFMC) and the temporo-parietal junction (TPJ) area for the control of shared representations and complex social-cognitive tasks, such as mental state attribution. In a functional magnetic resonance imaging experiment, we functionally dissociate the roles of TPJ and aFMC during the control of shared representations. Finally, we propose a hypothesis stating that the control of shared representations might be the missing link between functions of the mirror system and mental state attribution.

Keywords: imitation; inhibition; prefrontal cortex; temporo-parietal junction; mentalizing

1. INTRODUCTION

Converging evidence from cognitive psychology and neuroscience strongly suggests that perception and execution of action are tightly linked. The mere observation of an action activates a corresponding motor representation in the observer, suggesting that perception and action rely on a 'shared representational system' (Prinz 1997). Substantial efforts have been made to investigate the functional role of shared representations in imitation (Brass & Heyes 2005), action understanding (Rizzolatti & Craighero 2004; Hurley 2008) and also in empathy (Gallese 2003; Bastiaansen *et al.* 2009). Despite the efforts that have been devoted to investigate shared representations *per se*, one fundamental problem of a shared representational system has been widely neglected, namely how it is possible that such a system is able to distinguish between motor representations that have been internally generated and those that have been triggered by observing others' actions (Jeannerod 1999). This pivotal question of why we do not imitate all the time is central to the present article.

We will first briefly summarize evidence for shared representations of perception and action. Then we will introduce the idea that a shared representational system necessitates key processes related to the control of shared representations. We will then present previous evidence indicating that the control of shared representations involves processes that are also required for more complex social-cognitive skills such as mentalizing (i.e. our ability to infer other people's mental states). Support for this idea will be provided by a meta-analysis focusing on overlap of neuroimaging activations in two brain regions, the anterior fronto-median cortex (aFMC) and the temporo-parietal junction (TPJ). Furthermore, a functional magnetic resonance imaging (fMRI) study will be presented differentiating the functional roles of these two brain areas. Finally, we will propose an account that suggests that the control of shared representations provides the missing link between shared representations and social cognition.

(a) *Evidence for shared representations*

There is evidence from different research fields for the assumption that perception and execution of an action have a common representational basis (Brass & Heyes 2005). First, it has been demonstrated that the observation of an action primes the corresponding motor representation in the observer (Blakemore & Frith 2005; Massen & Prinz 2009). For example, executing an action while concurrently observing an incongruent action leads to slower responses compared with observation of a congruent action during action execution. Research on such 'motor priming' effects has focused on the conditions under which it occurs (Liepelt *et al.* 2008), the role of low-level factors in motor

priming (Bertenthal *et al.* 2006) and the influence of motor learning (Catmur *et al.* 2009). A second line of evidence for shared representations is provided by brain imaging studies showing that the observation of an action activates brain areas that are also involved in motor planning and execution (Grezes & Decety 2001). Research in this domain was initially motivated by the discovery of 'mirror neurons' in macaque monkeys, which were found to be active both when the monkey observes and executes a specific behaviour (Gallese *et al.* 1996). Brain imaging research thus focused primarily on defining the neural correlates of shared representations in humans and tried to reveal how these neural circuits relate to brain areas that have been demonstrated to house mirror neurons in monkeys. Furthermore, social psychology has started to investigate the role of shared representations in social cognition (Chartrand & Bargh 1999). This research focuses on the social consequences of imitation and under which conditions social imitation occurs (van Baaren *et al.* 2009). Finally, clinical neuropsychological research revealed that prefrontal patients sometimes display imitative response tendencies (Lhermitte *et al.* 1986; de Renzi *et al.* 1996; Brass *et al.* 2003). Patients with so-called 'imitation behaviour' tend to overtly imitate the experimenter, but when asked about the reason for the imitative behaviour, most patients claimed that they thought they were supposed to do so. Hence, imitation of the observed behaviour seems to turn into their intention, suggesting a problem in distinguishing self-generated motor intentions from externally triggered motor intentions.

To summarize, different strings of research support the idea of shared representations of perception and action. Furthermore, there is some indication that shared representations lead to automatic imitative response tendencies under specific contextual conditions and in patients with prefrontal lesions. However, research so far has widely neglected the precise functional mechanisms and brain circuits that are involved in the *control* of shared representations. In particular, the question arises how we can distinguish self-generated and externally triggered motor representations?

(b) *Is it me or is it you? Shared representations and self–other distinction*

Taking a closer look at action control theories, which highlight the underlying mechanisms and the acquisition of shared representations of action, may be helpful to further understand the problem of self–other distinction in a shared representational system (Brass & Heyes 2005). The ideomotor theory, as an instantiation of such a model, assumes that the perceivable consequences of an action become associated through learning with the motor programme. Motor representations hence contain the sensory consequences of actions, so that perceiving an action automatically activates the equivalent motor representation in the observer (Prinz 2005). As outlined earlier, this constitutes the basis for a shared representational system of perceived and internally planned actions ('common coding theory'; Prinz 1997). However, this shared

representational system does not explicitly code by whom the corresponding motor representation was caused (Jeannerod 1999). Hence, mechanisms are needed to keep apart self- and other-related motor representations (Decety & Grezes 2006). Such mechanisms allow avoiding automatic imitative behaviour in situations where an observed action leads to the activation of a motor representation that is not intended.

(c) *A common network for the inhibition of imitative behaviour, self–other distinction and mind reading*

In accordance with this assumption, recent brain imaging data suggest that the control of shared representations of action involves brain regions, which are related to self–other distinction and perspective taking (perspective taking can require taking a third person perspective in a bodily sense or taking the other person's mental perspective; Brass *et al.* 2001, 2005). Two fMRI studies investigated the neural correlates of controlling automatic imitative response tendencies, which can be used to index the control of shared representations. In this 'imitation-inhibition task', participants have to lift their index or middle finger in response to a number, while watching congruent (i.e. the same) or incongruent (i.e. the opposite) finger movements of a video-taped hand (Brass *et al.* 2000; figure 1*a*). In congruent trials, the video-taped hand executes a finger movement that is identical to the instructed movement; therefore, the participants are not required to distinguish their intended action from the observed action. Their response can be considered a 'quasi-imitative' reaction, which is normally indicated by faster reaction times on congruent trials (Brass *et al.* 2000). On the contrary, in incongruent trials, the instructed movement differs from the observed movement, which introduces an automatic, imitative tendency to execute the observed movement. This necessitates subjects to enforce the intended movement against the observed action, reflected in longer reaction times on incongruent trials (Brass *et al.* 2000). The required response on incongruent trials reflects therefore a non-imitative reaction with regard to the perceived movement. This allows measuring the interference effect (incongruent minus congruent condition for reaction times or errors). Controlling automatic imitative response tendencies yielded activation in the aFMC (Brass *et al.* 2001, 2005) and the TPJ (Brass *et al.* 2005). These brain regions have been implicated in the sense of agency (i.e. determine who is the cause or initiator of an action or thought; Farrer *et al.* 2003), perspective taking (Ruby & Decety 2003, 2004) and self-referential processing (Northoff & Bermpohl 2004). The aFMC and the TPJ are, in fact, the core brain regions of a network involved in mentalizing or having a 'theory of mind', defined as the ability to reflect on other people's mental states (e.g. Amodio & Frith 2006). This raises the interesting possibility that the control of shared representations involves similar functional-anatomical structures and cognitive mechanisms as more complex socio-cognitive skills ('functional overlap hypothesis').

(d) *Empirical evidence for a functional-anatomical overlap of imitation inhibition, self–other distinction and mentalizing*

To provide further evidence for the functional overlap hypothesis, we compared the results of the two imitation-inhibition studies within two meta-analyses. The first meta-analysis focused on the aFMC and included studies from two recent meta-analyses (Amodio & Frith 2006; Gilbert *et al.* 2006) and six additional Theory of Mind studies (ToM; Baron-Cohen *et al.* 1999; Gallagher *et al.* 2000; Rilling *et al.* 2004; den Ouden *et al.* 2005; Hynes *et al.* 2006; Mor-iguchi *et al.* 2006), resulting in a total of 19 studies on self-referential processing (22 activation peaks) and 26 studies on mentalizing (31 activation peaks). The second meta-analysis for the right TPJ was based on a recent meta-analysis (Decety & Lamm 2007) and four new studies on agency processing (Balslev *et al.* 2006; David *et al.* 2007; Tankersley *et al.* 2007; Farrer *et al.* 2008). Thus, in total, 19 studies (24 activation peaks) were included on agency tasks and 23 studies (28 activation maxima) on mentalizing. The results revealed a persuasive overlap of activations between imitative control and the social-cognitive functions in aFMC and TPJ (figure 1*b*).

However, while these meta-analyses provided evidence for a potential overlap of the control of shared representations and mentalizing across different studies, it is crucial to prove that such an overlap can be also demonstrated using a within-subject approach. Evidence for this claim was hitherto primarily based on descriptive, between-experiment and between-subject comparisons of activated regions. This makes interpretations more difficult owing to existing differences, for example, in imaging and analysis procedures between laboratories and differences in neuroanatomy between subject groups. Therefore, it would be more conclusive to observe this overlap of common, activated regions within one study and not between studies. We thus carried out a within-subject fMRI study, where we directly compared the neural circuits involved in the inhibition of imitative behaviour with brain circuits involved in social-cognitive processes (Spengler *et al.* in press). Participants completed a version of the imitation-inhibition task, a mentalizing task, a paradigm assessing self-referential judgments and agency processing. Activations in the individual tasks were tested for common overlap by means of a conjunction analysis. As predicted, commonly activated regions occurred selectively in aFMC and TPJ. Controlling imitation recruited a region in aFMC, overlapping with activations during mentalizing and self-referential thoughts. In the TPJ an area overlapped between imitative control, mentalizing and agency processing. These results mirror the meta-analytic data, showing an overlap of activated brain regions in aFMC and TPJ between studies on imitative control and social cognition.

Finally, we wanted to test whether deficits in the imitation-inhibition task are associated with impairments in perspective-taking and mentalizing. In order to do so, we correlated these tasks in brain-damaged patients with lesions around the TPJ and the prefrontal cortex (Spengler *et al.* submitted). If the inhibition of imitative behaviour shares neural resources with mentalizing and perspective taking these tasks should be strongly correlated in patients that vary regarding the integrity of these neural resources. Supporting the hypothesis, a highly significant correlation could be found between a mentalizing task and the imitation-inhibition task in the group with frontal lesions. Temporo-parietal lesioned patients showed a highly significant correlation between the imitation-inhibition task and both cognitive and visual perspective taking. Even after controlling for the performance in the control condition of the ToM task and several tests on executive functions, assessing response inhibition, mental flexibility and working memory, the results remained significant.

To summarize, our data strongly suggest that the inhibition of imitative behaviour overlaps with higher level social-cognitive abilities both at the functional and the neural level. However, while the research we reported focused primarily on the common role of aFMC and TPJ in the control of shared representations, there is strong evidence that their contribution to this function can be functionally dissociated.

(e) *Dissociating the role of anterior fronto-median cortex and temporo-parietal junction*

Despite extensive research on the role of TPJ and aFMC in social cognition, the differential role of these two areas is still poorly understood. The TPJ has been related to key computations in the social domain, such as agency processing and motor perspective taking (Ruby & Decety 2001; Blanke *et al.* 2002; Farrer & Frith 2002; Farrer *et al.* 2003), but also in other non-social processes, such as spatial attention (Corbetta *et al.* 2000; Mitchell 2008). On the other hand, this area is very often activated during mentalizing tasks (Saxe & Wexler 2005; Legrand & Ruby 2009). Thus, it is still an open question whether the TPJ comprises a single brain area or whether different subareas can be dissociated within the TPJ (Mitchell 2008).

The aFMC has similarly been related to a number of cognitive operations. Among the most dominant views is its role in self-referential processing and mentalizing (for an overview, see Amodio & Frith 2006; Legrand & Ruby 2009). Furthermore, the aFMC is part of the 'default network', a circuit of brain regions that show elevated activation levels in resting state situations (Raichle *et al.* 2001). We think that specifying the role of these two brain regions in the control of shared representations will allow us to gain deeper insight regarding their functional role in social cognition.

In the imitation-inhibition task TPJ and aFMC show a very similar activation pattern, with stronger activation in incongruent compared with congruent trials (Brass *et al.* 2005). Nevertheless, from a functional-anatomical point of view, it is very unlikely that both areas serve a similar function in the inhibition of imitative response tendencies. We have argued previously that the TPJ is involved in self–other distinction by indicating that the observed
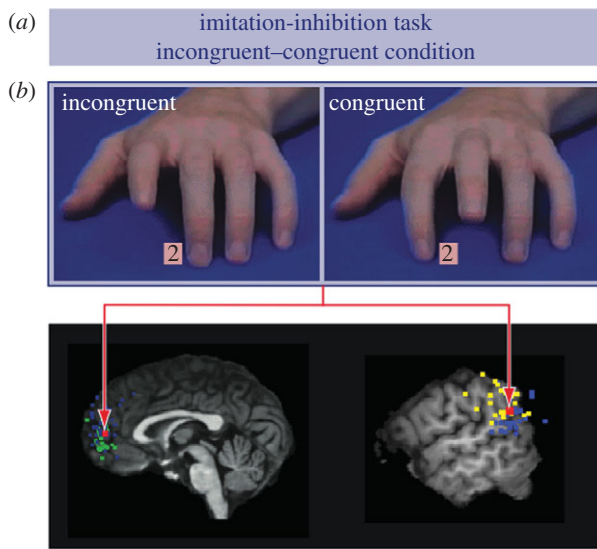
Figure 1. (*a*) Two frames depicted from the imitation-inhibition task. Participants have to respond to a number presented between the index and a middle finger of a video-taped hand while observing congruent or incongruent finger movements. (*b*) Brain activation for the contrast of incongruent versus congruent movements (Brass *et al.* 2005) plotted in red on a meta-analysis of the TPJ and the aFMC (blue, mentalizing; green, self-referential processing; yellow, agency processing). (For display purposes the activations of the social cognitive tasks are plotted with the same *x*-coordinate as the activations from the imitation-inhibition task.)

behaviour is related to another agent. Here, it is crucial to note that self–other confusion in the imitation-inhibition task presumably only occurs at a very early processing stage but not at the conscious level. Participants are always aware that the hand on the computer screen is not their hand. However, in such an early processing stage, attribution of the observed behaviour to the self might only occur if the observed behaviour is congruent to the planned behaviour, whereas observing an incongruent movement leads to an attribution of the observed behaviour to another person. In contrast, we assume that the aFMC is required to enforce one's own motor intention against the externally triggered response tendency. This is required when the observed action conflicts with the to-be-executed action. As long as participants observe the behaviour while they are preparing their response, these two cognitive operations of TPJ and aFMC are not dissociable because both are more strongly required in the incongruent condition. However, one can dissociate these two functions by changing the timing of the observed and instructed movement. When the observed behaviour follows execution of the instructed behaviour, different predictions result for TPJ and aFMC. TPJ should show stronger activation regardless of whether the observed movement is incongruent or congruent because a movement that is considerably delayed, with respect to the implementation of a motor intention, necessarily belongs to someone else. Therefore, even the congruent movement is, in these cases, attributed to another person. On the contrary, the aFMC should show no

activation regardless of congruency because participants have already implemented their motor intention before the congruent or incongruent movement occurs. Hence, presenting the congruent and incongruent finger movements after participants have executed their response should lead to an abolishment of the congruency effect in TPJ and aFMC. However, in TPJ, the activation level in the delayed condition should be similar to the incongruent condition of the simultaneous presentation (i.e. a relatively high level of activation), while in aFMC it should be similar to the congruent condition (i.e. a relatively low level of activation).

In a recent fMRI study, we tried to test these predictions. The basic logic of this experiment was based on the imitation-inhibition task that has been described earlier. Similarly, in the present paradigm, participants were required to execute index and middle finger movements and observed congruent or incongruent movements of the index or middle finger. However, we changed the original paradigm in two important ways. First, we administered a condition where participants either had to imitate or counter-imitate an index or a middle finger movement. Whether they had to imitate or counter-imitate was indicated by a symbolic stimulus. We called this condition the *simultaneous condition* because participants had to represent the observed movement and the instruction at the same time. The aim of this manipulation was to test whether we can replicate previous findings of aFMC and TPJ involvement with a paradigm that requires participants to respond to the observed movement in an imitative or non-imitative manner. Furthermore, we wanted to use this condition to localize the brain areas that are involved in overcoming imitative behaviour. The second and most crucial manipulation of the current experiment was related to the time at which participants observed the finger movements. In the *delayed condition*, participants first had to execute a lifting movement of the index or middle finger in response to a number. The observed finger movement was then triggered by the execution of the subject's finger movement and therefore only occurred after participants had executed their finger movement.

2. SUBJECTS AND METHODS

We tested 20 subjects with functional MRI. In the simultaneous condition, participants had to imitate the observed movement when a green cross appeared or had to counter-imitate when a red cross appeared. In the delayed condition, they had to lift their index finger from an infrared-light key when a '1' appeared and had to lift their middle finger when a '2' appeared. The offset of participant's finger from the response key triggered the presentation of the congruent or incongruent finger movement (figure 2). The simultaneous and delayed conditions were presented in separate blocks of 18 trials with 14 experimental trials (seven congruent and seven incongruent) randomly intermixed with four null events (a resting baseline in an event-related design). Each block was presented five times, resulting in 35 trials for each of the four conditions MAPPING (congruent versus incongruent) × DELAY
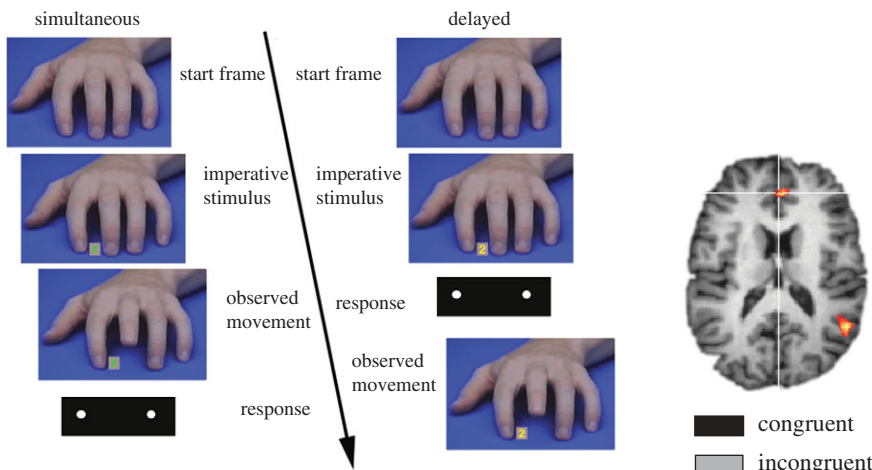
Figure 2. Schematic drawing of the modified imitation-inhibition task. The left side depicts the series of events in the *simultaneous* condition where participants had to imitate the observed movement when a green cross appeared and had to counter-imitate when a red cross appeared. The right side illustrates the *delayed* condition where participants had to respond to a number and the response triggered a congruent or incongruent movement.
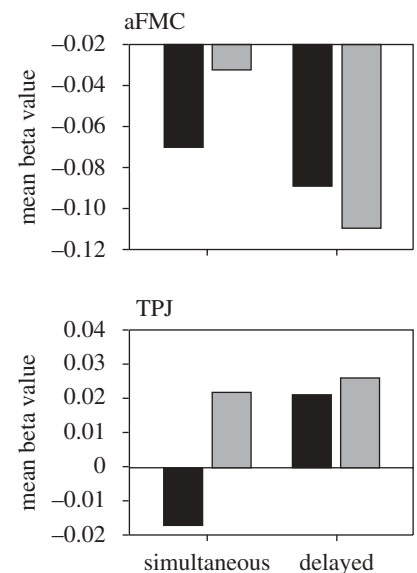

Figure 3. The left part shows the brain activation of incongruent versus congruent trials in the simultaneous condition. On the right, the signal strength analysis in TPJ and aFMC is displayed. The mean beta values in these two areas are plotted as a function of MAPPING (congruent versus incongruent) and DELAY (simultaneous, delayed).

(simultaneous versus delayed). The blocks were randomly intermixed with two other types of blocks (involving a motor imagery manipulation) that are not relevant for the current analysis.

Scanning was carried out on a 3T Bruker scanner with 20 axial slices (19.2 cm field of view; 64×64 matrix; 4 mm) using a standard EPI sequence (repetition time: 2000 ms; echo time: 30 ms; 90° flip angle). Data were analysed using the Lipsia software package. Data were movement and slicetime corrected. Then a spatial Gaussian filter with sigma = 1 and a temporal filter of 1/80 Hz was applied. All functional datasets were individually registered into the three-dimensional space using the participants' individual high resolution dataset. Finally, all data were interpolated to $3 \times 3 \times 3$ spatial resolution and linearly normalized to Talairach space. Data were modelled with a haemodynamic response function with a variable delay. The statistical analysis was carried out using a random-effects model.

First, we contrasted the incongruent and congruent condition of the simultaneous blocks to identify the brain areas that are involved in overcoming imitative behaviour. In accordance with previous work (Brass *et al.* 2005), we could identify significant brain activation ($p < 0.001$) in aFMC (x: 1, y: 39, z: 18) and in TPJ (x: 52, y: -54, z: 21; figure 3, left panel). In order to investigate the activation pattern of these brain regions in the delayed condition, we carried out a signal strength analysis in these two brain regions. We extracted the mean beta value from these coordinates for each experimental condition and entered them into a three-way ANOVA with the factors AREA (aFMC, TPJ), MAPPING (congruent versus incongruent) and DELAY (simultaneous, delayed). The behavioural data were also analysed with an ANOVA including only two factors (MAPPING, DELAY).

3. RESULTS

The behavioural results revealed a main effect for DELAY with faster responses in the delayed than in the simultaneous condition, $F(1, 19) = 143.5$, $p < 0.01$. Furthermore, a main effect of MAPPING was found, $F(1, 19) = 7.7$, $p < 0.05$. Participants were faster in the congruent than in the incongruent mapping. The interaction failed to reach significance, $F(1, 19) = 2.9$, $p = 0.10$. However, because we had an *a priori* hypothesis regarding differential MAPPING effects in the simultaneous and delayed condition (in fact a behavioural MAPPING effect in the delayed condition is logically impossible), we computed the MAPPING effect separately. While in the simultaneous condition reaction times were significantly faster in congruent (633 ms) compared with incongruent trials (670 ms), $t(19) = 2.5$, $p < 0.05$, in the delayed condition, no such effect was found (congruent: 478, incongruent: 484), $t(19) = 0.9$, $p = 0.38$. The MAPPING effect in the simultaneous condition indicates that participants indeed show interference similar to our previous studies, in which the observed movement was presented on the irrelevant dimension.

In a first step of the fMRI data analysis, we analysed aFMC and TPJ separately (figure 3, right panel). When splitting the analysis for the two brain regions the TPJ showed a DELAY by MAPPING interaction, $F(1, 19) = 5.1$, $p < 0.05$. A MAPPING effect (incongruent > congruent) was found for the simultaneous condition, $t(19) = 4.59$, $p < 0.001$, but was absent when the movement was presented after motor execution, $t(19) = 0.47$, $p = 0.64$. Furthermore, a MAPPING main effect was found, $F(1, 19) = 14.2$, $p < 0.01$, demonstrating that presentation of the movement after motor execution leads to a generally stronger activation of the TPJ.

It becomes obvious from figure 3 that the aFMC has negative beta values in all conditions. However, this is a common finding for this area that is difficult to interpret because it relies on the interpretation of the null line. Therefore, we will focus on relative differences between conditions. Similar to the TPJ, the aFMC also revealed a DELAY by MAPPING interaction, $F(1, 19) = 6.5$, $p < 0.05$. Participants showed a MAPPING effect (incongruent > congruent) in the simultaneous condition, $t(19) = 3.2$, $p < 0.01$, but no such mapping effect was found when the movement was shown after motor execution, $t(19) = -1.19$, $p = 0.24$. Furthermore, a main effect of DELAY showed a statistical trend, $F(1, 19) = 3.5$, $p = 0.076$. However, the direction of this effect was opposite to the TPJ with a stronger activation in the simultaneous condition compared with the condition where the movement was presented after motor execution.

Second, we entered both brain areas as a two-level factor in an additional ANOVA to test for interactions across brain regions. This analysis revealed a clear dissociation of aFMC and TPJ depending on whether the observed movement was presented simultaneously with the instruction (simultaneous condition) or after participants executed their response (delayed condition). This was reflected in the two-way interaction of REGION (TPJ, aFMC) and DELAY (simultaneous, delayed), $F(1, 19) = 4.7$, $p < 0.05$. While in the TPJ the activation for the delayed condition was stronger than for the simultaneous condition the opposite was true for the aFMC. In other words, the aFMC shows an elevated activity only when participants are involved in motor planning and see an incongruent movement. By contrast, the TPJ shows strong activity whenever the observed movement does not match the predicted outcome, namely when the observed movement is incongruent or delayed.

4. DISCUSSION

These data replicate our previous findings that overcoming interference from imitative behaviour involves the aFMC and the TPJ. However, the present design slightly differed from previous studies in which the observed behaviour was always irrelevant to the executed behaviour. In the simultaneous condition, participants had to select their behaviour on the basis of the observed behaviour (imitation or counter-imitation). In this sense, the congruent condition in the present design is an actual imitation condition. Furthermore, our data show that the congruency effect disappears when the movements are presented after motor execution. One might argue that this attenuation of the congruency effect is simply owing to the fact that participants do not attend to the movement when it is presented after movement execution. However, the elevated activation level in the TPJ clearly contradicts this interpretation. If participants did not attend to the delayed movement, why should they show stronger activation in the TPJ? Another potential objection against the present design might be that the imperative stimuli for the simultaneous and the delayed condition are different. However, as we could replicate our previous results in the simultaneous condition and

the most relevant manipulation was implemented in the delayed condition, we do not consider this point as very critical for the interpretation of our data.

Most importantly, however, these data strongly suggest that TPJ and aFMC serve different roles in the control of shared representations. Activation of TPJ seems to reflect the outcome of an agency judgement, namely whether the perceived sensory event is related to another agent or not (Farrer & Frith 2002; Farrer *et al.* 2003). By contrast, activation in aFMC is elevated only in the simultaneous condition where the motor plan conflicts with the observed behaviour. This can be explained by the involvement of the aFMC in managing the conflict between intended and externally triggered motor representations by enforcing one's own intention. This interpretation is very consistent with the potential role of the aFMC in mentalizing, subserving the representation of intentional states, such as beliefs, desires and intentions for both self and other (Amodio & Frith 2006). The information provided by the TPJ may possibly also be needed in the aFMC to keep the self and other perspective separate. In more abstract situations of mental state attribution, this mechanism might be required to tag a mental state as belonging to someone else. This is consistent with our idea that elementary, domain-general computations are needed during the inhibition of imitation as well as in social cognition (for a similar view, see Decety & Lamm 2007).

(a) The 'different-from-me' hypothesis of social cognition

The research we have presented so far indicates that two brain areas that can be functionally dissociated are involved in the control of shared representations and in higher order social-cognitive skills, such as mentalizing. But how does this functional-anatomical overlap relate to the proposal that shared representations form the basis for social cognition?

The 'like-me hypothesis' of social cognition (Meltzoff & Decety 2003) says that the experience that someone is 'like me' provides the basis for action understanding and mentalizing. This position suggests that the shared representational system (including the premotor and inferior parietal cortex) plays a fundamental role in action understanding (Gallese *et al.* 1996; Rizzolatti *et al.* 1996), and more complex forms of mentalizing, if the goals and intentions are not transparent (Gallese & Goldman 1998). However, as outlined above, recent research has consistently identified a network of brain activations in mentalizing tasks that does not include areas with mirror properties (Frith & Frith 1999, 2003). Furthermore, the position that the mirror system is involved in mentalizing has also been criticized on the basis of theoretical arguments (Damasio & Meyer 2008; Dinstein *et al.* 2008).

But does this imply that shared representations and mentalizing are not related at all? Our data might provide the missing link between mentalizing and the shared representation system. We suggest that common computational processes subserve both the intentional control of shared representations and also later-developing social-cognitive capacities (Brass & Spengler 2008). Accordingly, this proposal complements and extends

the proposal that shared representations form a basis for action understanding and social cognition (Gallese & Goldman 1998; Rizzolatti & Craighero 2004). Shared representations may thus be seen as a first step for more advanced mind-reading abilities, as an initial estimation of the goals of the 'mirrored' person (Frith & Frith 2006). Mirroring the responses of others might be ideal to constantly track and monitor the changing actions and emotions of interaction partners, but this motor or emotional contagion does not always and unequivocally convey the cause for this action or emotion (Mitchell *et al.* 2006). Conversely, the formation of higher level mentalizing capacities may be based on the ability to form representations of mental states of others and distinguish the other-perspective from the self-perspective (Decety & Grezes 2006). Consequently, this conceptualization of mentalizing functions is closely related to the control of a shared representational system as it is indexed by the inhibition of imitation. In accordance with this assumption, self–other distinction is also pivotal to the shared representation model of mental state attribution (Hurley 2005, 2008). Here, mirroring of actions and thus similarity between self and other arises first, but subsequently a mechanism has to come into play, which allows us to detect distinctions between self- and other-related representations.

From an ontogentetic perspective, a simple mechanism of associating perceptual consequences of an action with the motor program would be sufficient to constitute a shared representational system (Keysers & Perrett 2004; Brass & Heyes 2005). In the beginning, this system would not differentiate between consequences in the environment that are produced by other agents or oneself. For example, seeing someone else's hand moving can then automatically and unconsciously activate the associated motor representation for this action in the observer. Therefore, shared representations may be considered to be the 'default state' of the sensorimotor system. In the following step, a sense of agency has to be developed by a monitoring system, which is relying on the learning mechanism and the observation that specific sensory events in the environment are contingent upon one's own actions, while others are not. This experience that other people are different from oneself would allow the development of a sense of self and agency in the motor domain, which would in subsequent development be needed in mental state attribution and could thus be seen as a first precondition for later-developing mental states attribution abilities (Rochat 1999). However, one has to point out that these mechanisms, while being necessary, are not sufficient to develop the ability for mentalizing. In addition, cognitive processes are required that allow the formation, representation and integration of abstract mental content. Regarding functional neuroanatomy, additional areas are needed for mental state attribution, such as the temporal pole and the precuneus (Gallagher & Frith 2003). Activation of the precuneus is typically also found in self-referential tasks (Northoff & Bermpohl 2004). Both areas have been associated with the retrieval of autobiographical memory (Fink *et al.* 1996; Legrand & Ruby 2009) and

may be needed to create a wider context for the material currently being processed (Gallagher & Frith 2003). These processes and associated regions are thus required during mentalizing, but not during imitative control, reflected in partly different cortical networks.

(b) *Implications for clinical research*

Implications of this approach can also be applied to the understanding of developmental disorders such as autism spectrum disorders (ASDs). Previous studies in ASDs reported impairments of mentalizing (for an overview, see Frith 2003), and also weaker activations of typical 'theory of mind areas' (aFMC and TPJ) in neuroimaging studies (e.g. Castelli *et al.* 2002). Conversely, a second line of research focused on the integrity of the mirror system in autistic patients and has related this to poor social abilities and deficits in imitative performance in ASDs (Williams *et al.* 2001, 2006; Dapretto *et al.* 2006). However, to date, this account is still debated. In contrast to this hypothesis, the view favoured by the current work would predict that autistic patients should have problems in the control of imitative behaviour rather then in imitation *per se*. Recent evidence revealed no deficit in goal-directed imitation in autistic children, which speaks against a global failure in the mirror system in ASDs (Hamilton *et al.* 2007). It might be therefore possible that the mirror system is not deficient in ASDs, but that this system is not influenced by regions which distinguish between the self and other agents (Frith 2003). Impairments of such a system could therefore lead to egocentrism, abnormalities in self-awareness and limitations in mentalizing as they can be found in autism (Frith & de Vignemont 2005). This would also predict that the control of imitation might be related to social abilities (e.g. performance in ToM tasks) in individuals with ASD. In line with this idea, a recent model has suggested that a route responsible for automatic imitation is not generally disturbed, but that the modulation of this route is deficient in ASDs (Hamilton 2008; Southgate & Hamilton 2008). This 'top-down modulation' account predicts that functions of aFMC and TPJ are crucial to regulate automatic mimicry. These specific areas might modulate activity in other cerebral regions and therefore our model would predict that the aFMC and TPJ should show abnormal processing in ASDs, rather than typical 'mirror regions'.

5. CONCLUSIONS

We have briefly summarized evidence for the assumption that perception and execution of action share a common representational basis. Furthermore, we argued that TPJ and aFMC play a crucial role in controlling shared representations, and in mentalizing. In an fMRI experiment, we could dissociate the role of aFMC and TPJ. Finally, it was proposed that the control of shared representations might provide the missing link between mentalizing and shared representations by providing key processes that are relevant for both domains. In particular, both the inhibition of imitative behaviour and mental state attribution require self/other distinction, implemented by

the TJP, and representing conflicting mental states, presumably related to the aFMC.

REFERENCES

Amodio, D. M. & Frith, C. D. 2006 Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* **7**, 268–277. (doi:10.1038/nrn1884)

Balslev, D., Nielsen, F. A., Lund, T. E., Law, I. & Paulson, O. B. 2006 Similar brain networks for detecting visuo-motor and visuo-proprioceptive synchrony. *NeuroImage* **31**, 308–312. (doi:10.1016/j.neuroimage.2005.11.037)

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A. & Williams, S. C. 1999 Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* **11**, 1891–1898. (doi:10.1046/j.1460-9568.1999.00621.x)

Bastiaansen, J. A. C. J., Thioux, M. & Keysers, C. 2009 Evidence for mirror systems in emotions. *Phil. Trans. R. Soc. B* **364**, 2391–2404. (doi:10.1098/rstb.2009.0058)

Bertenthal, B. I., Longo, M. R. & Kosobud, A. 2006 Imitative response tendencies following observation of intransitive actions. *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 210–225. (doi:10.1037/0096-1523.32.2.210)

Blakemore, S. J. & Frith, C. 2005 The role of motor contagion in the prediction of action. *Neuropsychologia* **43**, 260–267. (doi:10.1016/j.neuropsychologia.2004.11.012)

Blanke, O., Ortigue, S., Landis, T. & Seeck, M. 2002 Stimulating illusory own-body perceptions. *Nature* **419**, 269–270. (doi:10.1038/419269a)

Brass, M. & Heyes, C. 2005 Imitation: is cognitive neuroscience solving the correspondence problem? *Trends Cogn. Sci.* **9**, 489–495. (doi:10.1016/j.tics.2005.08.007)

Brass, M. & Spengler, S. 2008 The inhibition of imitative behavior and attribution of mental states. In *Social cognition: development, neuroscience and autism* (eds T. Striano & V. Reid), pp. 52–66. Oxford, UK: Blackwell.

Brass, M., Bekkering, H., Wohlschlager, A. & Prinz, W. 2000 Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. *Brain Cogn.* **44**, 124–143. (doi:10.1006/brcg.2000.1225)

Brass, M., Zysset, S. & von Cramon, D. Y. 2001 The inhibition of imitative response tendencies. *NeuroImage* **14**, 1416–1423. (doi:10.1006/nimg.2001.0944)

Brass, M., Derrfuss, J., Matthes-von Cramon, G. & von Cramon, D. Y. 2003 Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology* **17**, 265–271. (doi:10.1037/0894-4105.17.2.265)

Brass, M., Derrfuss, J. & von Cramon, D. Y. 2005 The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia* **43**, 89–98. (doi:10.1016/j.neuropsychologia.2004.06.018)

Castelli, F., Frith, C., Happe, F. & Frith, U. 2002 Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125**, 1839–1849. (doi:10.1093/brain/awf189)

Catmur, C., Walsh, W. & Heyes, C. 2009 Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Phil. Trans. R. Soc. B* **364**, 2369–2380. (doi:10.1098/rstb.2009.0048)

Chartrand, T. L. & Bargh, J. A. 1999 The chameleon effect: the perception-behavior link and social interaction. *J. Pers. Soc. Psychol.* **76**, 893–910. (doi:10.1037/0022-3514.76.6.893)

Corbetta, M., Kincade, J. M., Ollinger, J. M., Mcavoy, M. P. & Shulman, G. L. 2000 Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* **3**, 292–297. (doi:10.1038/73009)

Damasio, A. & Meyer, K. 2008 Behind the looking-glass. *Nature* **454**, 167–168. (doi:10.1038/454167a)

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y. & Iacoboni, M. 2006 Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nat. Neurosci.* **9**, 28–30. (doi:10.1038/nn1611)

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R. & Vogeley, K. 2007 The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. *NeuroImage* **36**, 1004–1014. (doi:10.1016/j.neuroimage.2007.03.030)

Decety, J. & Grezes, J. 2006 The power of simulation: imagining one's own and other's behavior. *Brain Res.* **1079**, 4–14. (doi:10.1016/j.brainres.2005.12.115)

Decety, J. & Lamm, C. 2007 The role of the right tempero-parietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist* **13**, 580–593. (doi:10.1177/1073858407304654)

den Ouden, H. E., Frith, U., Frith, C. & Blakemore, S. J. 2005 Thinking about intentions. *NeuroImage* **28**, 787–796. (doi:10.1016/j.neuroimage.2005.05.001)

de Renzi, E., Cavalleri, F. & Facchini, S. 1996 Imitation and utilisation behaviour. *J. Neurol. Neurosurg. Psychiatry* **61**, 396–400. (doi:10.1136/jnnp.61.4.396)

Dinstein, I., Thomas, C., Behrmann, M. & Heeger, D. J. 2008 A mirror up to nature. *Curr. Biol.* **18**, R13–R18. (doi:10.1016/j.cub.2007.11.004)

Farrer, C. & Frith, C. D. 2002 Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage* **15**, 596–603. (doi:10.1006/nimg.2001.1009)

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J. & Jeannerod, M. 2003 Modulating the experience of agency: a positron emission tomography study. *NeuroImage* **18**, 324–333. (doi:10.1016/S1053-8119(02)00041-1)

Farrer, C., Frey, S. H., van Horn, J. D., Tunik, E., Turk, D., Inati, S. & Grafton, S. T. 2008 The angular gyrus computes action awareness representations. *Cereb. Cortex* **18**, 254–261. (doi:10.1093/cercor/bhm050)

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J. & Heiss, W. D. 1996 Cerebral epresentation of one's own past: neural networks involved in autobiographical memory. *J. Neurosci.* **16**, 4275–4282.

Frith, U. 2003 *Autism: explaining the enigma.* Oxford, UK: Blackwell Publishing.

Frith, U. & de Vignemont, F. 2005 Egocentrism, allocentrism, and Asperger syndrome. *Conscious. Cogn.* **14**, 719–738. (doi:10.1016/j.concog.2005.04.006)

Frith, C. D. & Frith, U. 1999 Interacting minds—a biological basis. *Science* **286**, 1692–1695. (doi:10.1126/science.286.5445.1692)

Frith, U. & Frith, C. D. 2003 Development and neurophysiology of mentalizing. *Phil. Trans. R. Soc. Lond. B* **358**, 459–473. (doi:10.1098/rstb.2002.1218)

Frith, C. D. & Frith, U. 2006 The neural basis of mentalizing. *Neuron* **50**, 531–534. (doi:10.1016/j.neuron.2006.05.001)

Gallagher, H. L. & Frith, C. D. 2003 Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* **7**, 77–83. (doi:10.1016/S1364-6613(02)00025-6)

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U. & Frith, C. D. 2000 Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* **38**, 11–21. (doi:10.1016/S0028-3932(99)00053-6)

Gallese, V. 2003 The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* **36**, 171–180. (doi:10.1159/000072786)

Gallese, V. & Goldman, A. 1998 Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2**, 493–501. (doi:10.1016/S1364-6613(98)01262-5)

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. 1996 Action recognition in the premotor cortex. *Brain* **119**, 593–609. (doi:10.1093/brain/119.2.593)

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D. & Burgess, P. W. 2006 Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *J. Cogn. Neurosci.* **18**, 932–948. (doi:10.1162/jocn.2006.18.6.932)

Grezes, J. & Decety, J. 2001 Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* **12**, 1–19. (doi:10.1002/1097-0193(200101)12:1<1::AID-HBM10>3.0.CO;2-V)

Hamilton, A. F. 2008 Emulation and mimicry for social interaction: a theoretical approach to imitation in autism. *Quart. J. Exp. Psychol.* **61**, 101–115. (doi:10.1080/17470210701508798)

Hamilton, A. F., Brindley, R. M. & Frith, U. 2007 Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia* **45**, 1859–1868. (doi:10.1016/j.neuropsychologia.2006.11.022)

Hurley, S. 2005 The shared circuits hypothesis: a unified functional architecture for control, imitation, and simulation. In *Perspectives on imitation* (eds S. Hurley & N. Chater). Cambridge, MA: MIT Press.

Hurley, S. 2008 The shared circuits model (SCM): how control, mirroring, and simulation can enable imitation, deliberation, and mindreading. *Behav. Brain Sci.* **31**, 1–22. discussion 22–58.

Hynes, C. A., Baird, A. A. & Grafton, S. T. 2006 Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* **44**, 374–383. (doi:10.1016/j.neuropsychologia.2005.06.011)

Jeannerod, M. 1999 The 25th Bartlett Lecture. To act or not to act: perspectives on the representation of actions. *Quart. J. Exp. Psychol. A* **52**, 1–29. (doi:10.1080/027249899391205)

Keysers, C. & Perrett, D. I. 2004 Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* **8**, 501–507. (doi:10.1016/j.tics.2004.09.005)

Legrand, D. & Ruby, P. 2009 What is self-specific? Theoretical investigation and critical review of neuroimaging results. *Psychol. Rev.* **116**, 252–282.

Lhermitte, F., Pillon, B. & Serdaru, M. 1986 Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: a neuropsychological study of 75 patients. *Ann. Neurol.* **19**, 326–334. (doi:10.1002/ana.410190404)

Liepelt, R., von Cramon, D. Y. & Brass, M. 2008 What is matched in direct matching? Intention attribution modulates motor priming. *J. Exp. Psychol. Hum. Percept. Perf.* **34**, 578–591. (doi:10.1037/0096-1523.34.3.578)

Massen, C. & Prinz, W. 2009 Movements, actions and tool-use actions: an ideomotor approach to imitation. *Phil. Trans. R. Soc. B* **364**, 2349–2358. (doi:10.1098/rstb.2009.0059)

Meltzoff, A. N. & Decety, J. 2003 What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Phil. Trans. R. Soc. Lond. B* **358**, 491–500. (doi:10.1098/rstb.2002.1261)

Mitchell, J. P. 2008 Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb. Cortex* **18**, 262–271. (doi:10.1093/cercor/bhm051)

Mitchell, J. P., Macrae, C. N. & Banaji, M. R. 2006 Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* **50**, 655–663. (doi:10.1016/j.neuron.2006.03.040)

Moriguchi, Y., Ohnishi, T., Lane, R. D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H. & Komaki, G. 2006 Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *NeuroImage* **32**, 1472–1482. (doi:10.1016/j.neuroimage.2006.04.186)

Northoff, G. & Bermpohl, F. 2004 Cortical midline structures and the self. *Trends Cogn. Sci.* **8**, 102–107. (doi:10.1016/j.tics.2004.01.004)

Prinz, W. 1997 Perception and action planning. *Eur J. Cogn. Psychol.* **9**, 129–154. (doi:10.1080/713752551)

Prinz, W. 2005 An ideomotor approach to imitation. In *Perspectives on imitation* (eds S. Hurley & N. Charter), pp. 141–156. Cambridge, MA: MIT Press.

Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A. & Shulman, G. L. 2001 A default mode of brain function. *Proc. Natl Acad. Sci. USA* **98**, 676–682. (doi:10.1073/pnas.98.2.676)

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. 2004 The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* **22**, 1694–1703. (doi:10.1016/j.neuroimage.2004.04.015)

Rizzolatti, G. & Craighero, L. 2004 The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. 1996 Premotor cortex and the recognition of motor actions. *Brain Res. Cogn. Brain Res.* **3**, 131–141. (doi:10.1016/0926-6410(95)00038-0)

Rochat, P. 1999 *Early social cognition: understanding others in the first months of life.* Mahawah, NJ: Lawrence Erlbaum Associates.

Ruby, P. & Decety, J. 2001 Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* **4**, 546–550.

Ruby, P. & Decety, J. 2003 What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur. J. Neurosci.* **17**, 2475–2480. (doi:10.1046/j.1460-9568.2003.02673.x)

Ruby, P. & Decety, J. 2004 How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* **16**, 988–999. (doi:10.1162/0898929041502661)

Saxe, R. & Wexler, A. 2005 Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* **43**, 1391–1399. (doi:10.1016/j.neuropsychologia.2005.02.013)

Southgate, V. & Hamilton, A. F. 2008 Unbroken mirrors: challenging a theory of autism. *Trends Cogn. Sci.* **12**, 225–229. (doi:10.1016/j.tics.2008.03.005)

Spengler, S., von Cramon, D. Y. & Brass, M. In press. Control of shared representations recruits key processes implicated in social cognition. *Hum. Brain Mapp.*

Spengler, S., von Cramon, D. Y. & Brass, M. Submitted. Intentional control of imitation is linked to aspects of social cognitive processing: neuropsychological evidence from prefrontal and temporo-parietal lesions.

Tankersley, D., Stowe, C. J. & Huettel, S. A. 2007 Altruism is associated with an increased neural response to agency. *Nat. Neurosci.* **10**, 150–151. (doi:10.1038/nn1833)

van Baaren, R., Janssen, L., Chartrand, T. L. & Dijksterhuis, A. 2009 Where is the love? The social aspects of mimicry. *Phil. Trans. R. Soc. B* **364**, 2381–2389. (doi:10.1098/rstb.2009.0057)

Williams, J. H., Whiten, A., Suddendorf, T. & Perrett, D. I. 2001 Imitation, mirror neurons and autism. *Neurosci. Biobehav. Rev.* **25**, 287–295. (doi:10.1016/S0149-7634(01)00014-8)

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D. & Whiten, A. 2006 Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia* **44**, 610–621. (doi:10.1016/j.neuropsychologia.2005.06.010)