Functional Specificity and Sex Differences in the Neural Circuits Supporting the Inhibition of Automatic Imitation

Kohinoor M. Darda*, Emily E. Butler*, and Richard Ramsey

Abstract
Humans show an involuntary tendency to copy other people's actions. Although automatic imitation builds rapport and affiliation between individuals, we do not copy actions indiscriminately. Instead, copying behaviors are guided by a selection mechanism, which inhibits some actions and prioritizes others. To date, the neural underpinnings of the inhibition of automatic imitation and differences between the sexes in imitation control are not well understood. Previous studies involved small sample sizes and low statistical power, which produced mixed findings regarding the involvement of domain-general and domain-specific neural architectures. Here, we used data from Experiment 1 (N = 28) to perform a power analysis to determine the sample size required for Experiment 2 (N = 50; 80% power). Using independent functional localizers and an analysis pipeline that bolsters sensitivity, during imitation control we show clear engagement of the multiple-demand network (domain-general), but no sensitivity in the theory-of-mind network (domain-specific). Weaker effects were observed with regard to sex differences, suggesting that there are more similarities than differences between the sexes in terms of the neural systems engaged during imitation control. In summary, neurocognitive models of imitation require revision to reflect that the inhibition of imitation relies to a greater extent on a domain-general selection system rather than a domain-specific system that supports social cognition.

INTRODUCTION

Human social interactions are guided by nonverbal cues, such as copying behaviors. In the last two decades, much research has investigated the involuntary tendency to copy other’s actions—a phenomenon known as automatic imitation (Heyes, 2011). Automatic imitation is thought to be beneficial in social situations because it develops affiliative attitudes, better cooperation, and feelings of closeness between interacting partners (Chartrand & Lakin, 2013). Prior neuroscience research has shown that imitation is supported by the mirror neuron system, a neural network engaged in perceiving and performing actions (Iacoboni, 2009; Rizzolatti & Craighero, 2004; Iacoboni et al., 1999). Imitation, however, is unlikely to rely on a single cognitive or brain system (Southgate & Hamilton, 2008). For example, in many circumstances, imitation is maladaptive and requires inhibition (Cross & Iacoboni, 2014; Cross, Torrissi, Losin, & Iacoboni, 2013; van Schie, van Waterschoot, & Bekkering, 2008; Newman-Norlund, van Schie, van Zuijlen, & Bekkering, 2007). In such situations, a selection mechanism is required to suppress the tendency to imitate and prioritize alternative actions (Brass, Ruby, & Spengler, 2009). To date, studies investigating the neural mechanisms of imitation control have been limited by small sample sizes and low statistical power, which has produced mixed findings (Table 1). Furthermore, no neuroscience research has investigated how individual differences such as sex modulate imitation control, even though behavioral research has shown that imitative tendencies vary as a function of sex (Butler, Ward, & Ramsey, 2015; Sonnby-Borgström, Jönsson, & Svensson, 2008; Dimberg & Lundquist, 1990). Across two fMRI experiments, which had higher statistical power and functional sensitivity than prior studies, we investigated the extent to which imitation inhibition relies on a domain-specific or domain-general neural network, which varies its response as a function of sex.

Much like cognitive science in general (Kanwisher, 2010; Hirschfeld & Gelman, 1994), inhibitory control research has focused on a neat division between domain-general and domain-specific mental operations. Domain-general inhibitory systems, which operate across multiple tasks, have been identified in dorsal frontoparietal cortices (Aron, Robbins, & Poldrack, 2014; Hazeltine, Poldrack, & Gabrieli, 2007; Nee, Wager, & Jonides, 2007; Wager et al., 2005; Bunge, Hazeltine, Scanlon, & Rosen, 2002). This brain circuit has been labeled the multiple demand (MD) network because of its engagement in a diversity of mental operations (Duncan, 2010). By contrast, evidence from fMRI, neurostimulation, and neuropsychological patient studies has suggested that a domain-specific...
circuit in an anterior portion of medial prefrontal cortex (mPFC) and right temporo-parietal junction (rTPJ) operates during the inhibition of imitation (Bardi, Gheza, & Brass, 2017; Santiesteban, Banissy, Catmur, & Bird, 2012, 2015; Sowden & Catmur, 2015; Hogeveen et al., 2014; Klapper, Ramsey, Wibboldus, & Cross, 2014; Wang, Ramsey, & Hamilton, 2011; Spengler, von Cramon, & Brass, 2009, 2010; Brass, Derrfuss, Matthes-von Cramon, & von Cramon, 2003; Brass, Zysset, & von Cramon, 2001). Beyond the control of imitation, mPFC and rTPJ have been consistently implicated in a variety of social cognition functions, which require distinguishing between self and other, as well as reasoning about other people’s mental states (theory of mind [ToM]; Van Overwalle, 2009; Amodio & Frith, 2006; Saxe & Kanwisher, 2003; Frith & Frith, 1999). These results led to theorizing that a key neural circuit for social cognition also regulates imitative tendencies (Brass et al., 2009).

Although theories of imitation control have been developed that are based on functioning of the ToM network, evidence from fMRI studies that used an RT measure of imitation inhibition have not provided consistent support for the involvement of a domain-specific neural network (Table 1). The RT measure of imitation involves making finger movements while simultaneously watching compatible or incompatible finger movements (Brass, Bekkering, Wohlschlager, & Prinz, 2000; Stürmer et al., 2000). The difference between RTs in these two conditions (i.e., the general compatibility effect) has been argued to index imitative control, as greater cognitive resources are required to inhibit movements that are incompatible to one’s own responses (Heyes, 2011; Brass & Heyes, 2005). Approximately half of the fMRI studies using this paradigm failed to find engagement of rTPJ and anterior mPFC. In addition, a number of studies showed engagement of regions associated with the MD network, including dorsal fronto-parietal cortex, supplementary motor area (SMA) and anterior insula (Marsh, Bird, & Catmur, 2016; Cross & Iacoboni, 2013; Mengotti, Corradi-Dell’Acqua, & Rumiati, 2012; Crescentini, Mengotti, Grecucci, & Rumiati, 2012; Cross et al., 2013; Klapper et al., 2014; Wang et al., 2011; Spengler, von Cramon, & Brass, 2009, 2010; Brass, Derrfuss, Matthes-von Cramon, & von Cramon, 2003; Brass, Zysset, & von Cramon, 2001). Moreover, the most common measure of imitation interference is confounded by spatial compatibility or the tendency to respond faster to a stimulus when it is on the same side of space as the response (e.g., Simon, 1969). To measure imitation interference independent of spatial compatibility effects, spatial and imitative processes need to be dissociated (Gowen, Bolton, & Poliakoff, 2016; Marsh et al., 2016; Boyer, Longo, & Bertenthal, 2012; Cooper, Catmur, & Heyes, 2012; Catmur & Heyes, 2011; Wiggett, Hudson, Tipper, & Downing, 2011; Bertenthal, Longo, & Kosobud, 2006). Therefore, the extent to which imitation inhibition relies on domain-specific and domain-general architectures remains unclear. Indeed, no

Table 1. fMRI Studies Investigating Imitation Control Using Modified Versions of the Imitation Inhibition Task

<table>
<thead>
<tr>
<th>Sample (Male:Female)</th>
<th>Dissociation of Imitative and Spatial Processes</th>
<th>Analysis</th>
<th>Brain Networks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROI</td>
<td>Whole-brain</td>
<td>mPFC</td>
</tr>
<tr>
<td>Brass et al., 2001</td>
<td>10 (4:6)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Brass et al., 2005</td>
<td>20 (8:12)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Brass et al., 2009</td>
<td>20a</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Spengler et al., 2009</td>
<td>18 (9:9)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Bien et al., 2009</td>
<td>15 (5:10)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Crescentini et al., 2011</td>
<td>19 (9:10)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cross &amp; Iacoboni, 2013</td>
<td>24 (12:12)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mengotti et al., 2012</td>
<td>22 (10:12)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cross et al., 2013</td>
<td>25 (5:15)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Klapper et al., 2014</td>
<td>19 (2:17)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Marsh et al., 2016</td>
<td>24 (7:17)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Wang et al., 2011</td>
<td>20 (5:15)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Evidence that the engagement of mPFC and rTPJ is inconsistent across fMRI studies that investigated imitation control using modified versions of the imitation inhibition task. For all studies, engagement of mPFC or rTPJ is reported only for contrasts that test for inhibiting the urge to automatically imitate. Engagement of the MD network is reported only for whole-brain analyses. Except for Wang et al. (2011), which used hand movements, all other tasks used modified versions of the imitation inhibition tasks involving finger movements (Brass et al., 2000). For a more detailed version of this table, see Supplementary Table S6.

aNumber of male and female individuals not mentioned.

bMPFC showed engagement only at p < .005, uncorrected.
research to date has dissociated spatial from imitative processes and used a functional ROI (fROI) approach (Fedorenko, Duncan, & Kanwisher, 2013; Kanwisher, 2010). Using a fROI approach enables investigation of how functionally defined brain circuits, such as the MD and ToM networks, operate during the control of imitation.

A further area of imitation research that has received little attention is the extent to which imitative control varies across individuals, especially between the sexes. It has been argued that imitation is modulated by stable individual differences, such as empathy (Chartrand & Lakin, 2013) and sex (Butler et al., 2015; Sonnby-Borgström et al., 2008). Although it has been suggested that women excel across a range of social processes compared with men (Baron-Cohen, 2002), only a limited number of studies have investigated sex differences in social cognition, and the results are often mixed, do not replicate, or are specific to very select contexts or samples (Hyde, 2014; Miller & Halpern, 2014). Furthermore, studies of sex differences in social cognition have mainly focused on emotional expression perception and mental state reasoning with little emphasis placed on imitation (Krach et al., 2009; Russell, Tchanturia, Rahman, & Schmidt, 2007; Rahman, Wilson, & Abrahams, 2004; Campbell et al., 2002; Thayer & Johnsen, 2000).

A recent study that used an RT measure of imitation inhibition (Brass et al., 2000) showed that female individuals showed a greater level of interference than male individuals (Butler et al., 2015). It is possible that this sex difference in imitation control may be mediated by empathy—female individuals have been shown to be more empathetic compared with male individuals (Christov-Moore et al., 2014; Baron-Cohen & Wheelwright, 2004). However, even though empathy has been associated with different types of imitation paradigms (Müller, Leeuwen, Baaren, Belkering, & Dijksterhuis, 2013; Sonnby-Borgström, Jönsson, & Svensson, 2003; Sonnby-Borgström, 2002, 2010, 2017; Clayson, Clawson, & Larson, 2011; Rubia, Hyde, Halari, Giampietro, & Smith, 2010; Bayliss, di Pellegrino, & Tipper, 2005). It is possible, therefore, that a domain-general system may underpin the sex differences observed across these tasks, including during imitation control, but no research to date has directly investigated this proposal.

Across two fMRI experiments, the current study investigated functional specificity and sex differences in imitation control. Several aspects of the experimental design provide grounds to extend current understanding in meaningful and concrete ways. First, this is the first study to use independent functional localizers to identify MD and ToM networks in single subjects and directly test the involvement of these networks in imitation control. By doing so, we can directly test hypotheses regarding the role of functionally defined neural circuits (i.e., MD and ToM networks) and therefore minimize the reliance on reverse inference to infer cognitive function based on anatomical localization (Poldrack, 2006). Second, we used data from Experiment 1 to perform a power analysis to determine the sample size required to achieve a desired level of power in Experiment 2. Given the inconsistent findings in prior studies, which had relatively small sample sizes, this multieperiment approach made sure that our key experiment had over 80% power to detect expected effect sizes. Third, to avoid spatial compatibility confounds, in Experiment 2, we used a modified version of the imitation inhibition paradigm that allowed for an independent measure of spatial and imitative compatibility (Catmur & Heyes, 2011). If the inhibition of automatic imitation relies on a domain-specific neural architecture that is associated with social cognition, as proposed by Brass and colleagues (2009), mPFC and rTPJ would be engaged in imitation control. In contrast, engagement of the MD network would suggest that domain-general processes subserve imitation control. Furthermore, the sex difference found previously (Butler et al., 2015) may be supported by differences in ToM or MD networks.

METHODS
Overview of the Experimental Approach
Experiment 1 used a group-level whole-brain analysis, which provided the basis for power analyses that set up Experiment 2 as the critical experiment with high statistical power (80%). In Experiment 2, to increase sensitivity and functional resolution, we used independent localizers to identify key functional circuits (i.e., MD and ToM networks), and analyses were performed in single subjects to precisely quantify the consistency of

916 Journal of Cognitive Neuroscience Volume 30, Number 6

Downloaded from http://direct.mit.edu/jocn/article-pdf/30/6/914/1787529/jocn_a_01261.pdf by guest on 24 September 2021
network engagement across individuals (Nieto-Castañón & Fedorenko, 2012; Kanwisher, 2010). Group-level analyses require responses across individuals to overlap in individual voxels. In contrast, the fROI approach allows identification of corresponding functional regions without the requirement of exact voxel overlap across individuals. Therefore, the same voxels need not be active across individuals, as long as voxels within a functionally defined ROI are consistently active across individuals. Consequently, group-level analyses may underestimate functional specificity, whereas fROI analyses can show increased sensitivity (Nieto-Castañón & Fedorenko, 2012). In addition, because of a constrained search volume, fROI analyses typically have higher statistical power than whole-brain analyses (Fedorenko, Hsieh, Nieto-Castañón, Whitfield-Gabrieli, & Kanwisher, 2010; Saxe, Brett, & Kanwisher, 2006).

### Experiment 1

#### Participants

Twenty-eight participants ($M_{\text{age}} = 23.96$, $SD_{\text{age}} = 5.52$; 14 women) participated for monetary compensation of £15. Participants gave informed consent in line with the guidelines set by the Research Ethics and Governance Committee of the School of Psychology at Bangor University, were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological damage.

#### Design and Procedure

All participants performed the imitation task inside the scanner. The participants also did four additional tasks in the same scanning session as part of another experiment. The scanning session started with the imitation task, followed by a run of a face perception task, a flanker task (Erikson & Erikson, 1974), another run of the face perception task, a dynamic face localizer (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011), and a ToM localizer (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011). The order of the tasks was counterbalanced across participants such that, of the 28 participants, 14 participants did the imitation task first, and 14 participants did the flanker task first, with the order of the other tasks remaining the same.
The imitation inhibition task. The imitation task was based on a stimulus–response compatibility paradigm developed by Brass et al. (2000) consisting of observation and execution of finger-lifting movements during MRI scanning. Before the task, participants were instructed to hold down the “blue” and “yellow” buttons on the response box with their index and middle fingers of the right hand, respectively. A number cue (either “1” or “2”) was presented to participants, and they were asked to lift their index finger on presentation of the number “1” and the middle finger for the number “2.” Simultaneously, they also viewed an image of an index or middle finger lift of a left hand viewed from the third-person perspective, such that the fingers extended toward the participants. Thus, there were four trial types in an event-related design that led to two conditions—participants performing the same (congruent) or different (incongruent) finger movement to the observed hand image.

Each trial started with a fixation cross (500 msec) followed by a neutral hand (for a random ISI of 500, 700, or 1000 msec) and a hand image with an index/middle finger lift, which stayed onscreen for 2000 msec, irrespective of when the participant made the response. Sequencing the hand images in such a way led to the appearance of apparent motion of the finger. After 2000 msec, the next trial started immediately with a fixation cross (500 msec). To separately model the influence of individual events in an event-related design, the four trial types were pseudorandomized, such that each trial type was preceded by each other trial type and by itself an equal number of times (Wager & Nichols, 2003; Josephs & Henson, 1999). There were 17 trials in each block. The first trial was used to set up the randomization sequence but excluded from the analysis as it was not preceded by any other trial. The remaining 16 trials within a block were analyzed and consisted of eight trials per condition. Each run consisted of five blocks separated by a 3-, 4-, or 5-sec fixation cross. All participants completed one run of the imitation task. Thus, there were 80 trials of interest (40 congruent and 40 incongruent).

Behavioral Data Analysis

RT on the imitation inhibition task was measured as the time from number cue onset to when participants made a response. To ensure participants were engaging correctly with the task, participants who had less than 80% accuracy were removed. In addition, RTs more than 3 S.Ds away from the mean were excluded from the analyses. Furthermore, trials on which participants made an “error” were excluded from the analyses. Errors included an incorrect response, no response, a response after 2000 msec, and pressing an invalid key. The general compatibility effect was calculated as the RT difference between incompatible and compatible trials. A one-sample t test was performed to verify the presence of a general compatibility effect. A one-tailed independent sample t test was performed to determine if the compatibility effect was greater for female than male individuals. Mean differences, 95% confidence intervals (CI), and Cohen’s d (Cohen, 1992) are reported for all effects of interest. For the one-sample t test, Cohen’s d was calculated as mean difference divided by the standard deviation of the sample (Lakens, 2013). The 95% CI is reported for the lower bound for a one-tailed t test. For the independent samples t test, Cohen’s d was calculated as mean difference between the two groups divided by the pooled standard deviation (Cohen, 1992).

fMRI Data Analysis

Data acquisition. Participants were placed supine in a 3-T Philips MRI scanner using a SENSE 32-channel phased array coil. They were requested to avoid head motion during the scanning session and were presented stimuli on a computer screen placed behind the scanner made visible by a mirror attached to the head coil. Responses on the task were recorded with the help of a button box that recorded RTs. Thirty-five axial slices were acquired in an ascending order using a T2*-weighted EPI sequence. The reference slice for slice time correction was the slice acquired in the middle of the sequence (Slice 17). Parameters are as follows: voxel size = 3 × 3 × 4 mm, repetition time = 2000 msec, echo time = 30 msec, flip angle = 90°, slice thickness = 4 mm, slice gap = 0.8 mm, field of view = 230 × 230 × 167 mm³. One hundred seventy-four volumes were collected for the imitation task.

Four dummy scans collected at the beginning of each run of the task were not included in any analyses. A high-resolution T1-weighted anatomical image was also collected with the following parameters: repetition time = 12 msec, echo time = 3.5 msec, flip angle = 8°, number of axial slices = 170, voxel size = 1 mm³, field of view = 250 × 250 × 170 mm³.

Data preprocessing and general linear model. Functional images were preprocessed in SPM-8. Data were realigned, unwarped, and corrected for slice timing. Data were normalized to the Montreal Neurological Institute (MNI) template with a resolution of 3 mm³, and images were spatially smoothed (8 mm).

For the imitation task, a design matrix was fit for each participant with three regressors: one each for the correct trials of the two conditions, and one for the “new” trials (i.e., the first trial of each block). The new trials were not used in any further analyses. Stimulus onsets were time-locked to the presentation of the number cue with a duration of 0 sec and convolved with the standard hemodynamic response function.

Whole-brain analyses. Contrast images (incompatible > compatible) were calculated at the single-subject level for the imitation inhibition task to identify regions of the brain showing a compatibility effect. Group-level contrast images were created from these single-subject contrast images to identify regions that were consistently engaged for the compatibility effect across the sample using one-sample
tests. To identify a neural signature of the sex difference in imitation inhibition, a Sex × Compatibility ANOVA was computed (female [incompatible > compatible] > male [incompatible > compatible]) as female individuals have been shown to have a higher compatibility effect than male individuals in the imitation task (Butler et al., 2015). For all analyses, contrast images were taken to the group level and thresholded using a voxel-level threshold of \( p < .001 \) and a voxel extent of 10 voxels. Correction for multiple comparisons was performed at the cluster level (Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994), with clusters that survive correction for multiple corrections using a family-wise error correction (\( p < .05 \); shown in bold font in Table 2A and B; see Results). This restricts the likelihood of false positives (Eklund et al., 2016). Clusters of activity were identified with the SPM Anatomy toolbox (Eickhoff et al., 2005).

**Experiment 2**

**Participants**

Fifty-five participants (Mean age = 22.04, SD age = 3.70; 27 women) were recruited from the Bangor community and were either reimbursed with £15 or three course credits for their participation. Informed consent was obtained in line with the guidelines set by the Research Ethics and Governance Committee of the School of Psychology at Bangor University. All participants were right-handed, did not have dyslexia or dyspraxia, were not on any medication, did not report neurological damage, and had normal or corrected-to-normal vision. The sample size was determined by a power analysis based on Experiment 1 data (see Results).

**Design and Procedure**

Each participant performed three tasks inside the scanner—the automatic imitation task, a ToM network localizer task, and an MD network localizer task. The order of the tasks was as follows: two runs of the MD network localizer task were interspersed between three runs of the imitation task to offset boredom. This was followed by two runs of the ToM network localizer task. The ToM task was always presented at the end to reduce the likelihood that belief reasoning during the ToM task would influence performance in the imitation task. The order was the same for all participants. Participants also completed a 50-item International Personality Item Pool questionnaire (Donnellan, Oswald, Baird, & Lucas, 2006; Goldberg, 1992; unrelated to the current study) and a stimulus rating form where they were asked to rate the hand stimulus from the imitation task as either male, female, or neutral. The entire session lasted approximately 1.5 hr, with 60 min inside the scanner. All stimulus presentation was coded in MATLAB 2015b and presented with PsychToolBox 3.0.6.

The imitation inhibition task. The automatic imitation task was similar to the one used in Experiment 1, but with two changes. First, we used a different hand stimulus, which was rated as sex-neutral by observers. The sex of the hand was an important consideration to minimize the possibility of an own-sex bias while exploring sex differences in imitation inhibition. As such, we conducted pilot work that asked observers to evaluate a range of hand stimuli in terms of masculinity and femininity; and we selected the most sex-neutral stimulus (see Supplementary Information, Development of Stimuli). We only used one hand stimulus to simplify the design space. Although using one sex-neutral hand stimulus provided greater experimental control, it may have harmed our ability to study or elicit sex differences. Future work could probe this further by varying the sex of the stimulus and/or by using more sex-typical stimuli.

The second change that we made was to calculate an imitative compatibility effect independent of spatial compatibility (Catmur & Heyes, 2011). To do so, participants viewed an image of an index or middle finger lift of either a right or left hand but always responded with their right hand. Using right- and left-hand images produced eight trial types and four main conditions of interest (see Figure 2A). For example, when cued to lift their index finger while observing a left-hand index finger lift, the observed movement is both imitatively compatible (same finger), as well as spatially compatible (same side of space to the executed movement). In contrast, when observing a right-hand index finger lift, the participant’s response is imitatively compatible (same finger), but it is not on the same side of space (they are spatially incompatible). Thus, participants performed the same (imitatively compatible) or different (imitative incompatible) finger movement on the same (spatially compatible) or different (spatially incompatible) side of space to the observed finger movement, giving rise to the following four conditions:

1. Imitatively compatible and spatially compatible
2. Imitatively compatible and spatially incompatible
3. Imitatively compatible and spatially incompatible
4. Imitatively incompatible and spatially compatible

Sequencing information and pseudorandomization was the same as Experiment 1. There were 65 trials in each block. The first trial was used to set up the randomization sequence but excluded from the analysis as it was not preceded by any other trial. The remaining 64 trials were analyzed, consisting of 16 trials per condition. Each run consisted of two blocks separated by a 3-sec fixation cross. All participants completed three runs of the imitation task. In total, there were 384 trials of interest, 96 per condition. Experiment 2, therefore, had more than twice the number of trials per condition than Experiment 1.

**Localizer tasks**

The MD network localizer. To identify regions of the MD network, a verbal working memory (WM) task was used (Fedorenko, Behr, & Kanswisher, 2011). Participants
Table 2. General Compatibility Effect and Sex × Compatibility Interaction for the Imitation Inhibition Task (Experiment 1)

<table>
<thead>
<tr>
<th>Region</th>
<th>Cluster Size</th>
<th>p FWE Corr</th>
<th>t</th>
<th>x</th>
<th>y</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) General Compatibility Effect (Incompatible &gt; Compatible)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L IPL extending into SPL and superior frontal gyrus</td>
<td>986</td>
<td>.&lt;.001</td>
<td>8.40</td>
<td>−39</td>
<td>−40</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.50</td>
<td>−36</td>
<td>−37</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.38</td>
<td>−27</td>
<td>−7</td>
<td>70</td>
</tr>
<tr>
<td>L cerebellum</td>
<td>150</td>
<td>.001</td>
<td>5.79</td>
<td>−21</td>
<td>−55</td>
<td>−41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.95</td>
<td>−30</td>
<td>−55</td>
<td>−35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.72</td>
<td>−9</td>
<td>−70</td>
<td>−44</td>
</tr>
<tr>
<td>R cerebellum</td>
<td>198</td>
<td>.&lt;.001</td>
<td>5.71</td>
<td>21</td>
<td>−58</td>
<td>−44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.12</td>
<td>45</td>
<td>−46</td>
<td>−32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.32</td>
<td>39</td>
<td>−55</td>
<td>−23</td>
</tr>
<tr>
<td>R PrecG extending across superior frontal gyrus and MFG</td>
<td>183</td>
<td>.&lt;.001</td>
<td>5.12</td>
<td>27</td>
<td>−1</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.04</td>
<td>42</td>
<td>2</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.39</td>
<td>39</td>
<td>−10</td>
<td>61</td>
</tr>
<tr>
<td>R postcentral gyrus extending into SPL and IPL</td>
<td>481</td>
<td>.&lt;.001</td>
<td>5.18</td>
<td>33</td>
<td>−40</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.55</td>
<td>42</td>
<td>−40</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.50</td>
<td>48</td>
<td>−34</td>
<td>37</td>
</tr>
<tr>
<td>R posterior middle temporal gyrus</td>
<td>41</td>
<td>.179</td>
<td>4.61</td>
<td>66</td>
<td>−46</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.13</td>
<td>57</td>
<td>−40</td>
<td>−8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.63</td>
<td>60</td>
<td>−43</td>
<td>13</td>
</tr>
<tr>
<td>L insula</td>
<td>24</td>
<td>.458</td>
<td>4.67</td>
<td>−36</td>
<td>17</td>
<td>−2</td>
</tr>
<tr>
<td>R posterior medial frontal cortex</td>
<td>20</td>
<td>.564</td>
<td>4.79</td>
<td>3</td>
<td>−4</td>
<td>73</td>
</tr>
<tr>
<td>L posterior medial frontal cortex</td>
<td>55</td>
<td>.083</td>
<td>4.40</td>
<td>−3</td>
<td>−1</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.82</td>
<td>−6</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>R pallidum extending into thalamus</td>
<td>11</td>
<td>.834</td>
<td>4.14</td>
<td>21</td>
<td>−7</td>
<td>−2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.80</td>
<td>15</td>
<td>−6</td>
<td>7</td>
</tr>
<tr>
<td>L paracentral lobule</td>
<td>11</td>
<td>.834</td>
<td>3.89</td>
<td>−12</td>
<td>−19</td>
<td>79</td>
</tr>
<tr>
<td>R middle cingulate cortex</td>
<td>20</td>
<td>.564</td>
<td>3.85</td>
<td>9</td>
<td>14</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.78</td>
<td>6</td>
<td>8</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.67</td>
<td>6</td>
<td>17</td>
<td>52</td>
</tr>
<tr>
<td>(B) Sex × Compatibility (Female (Incompatible &gt; Compatible) &gt; Male (Incompatible &gt; Compatible))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L SPL extending into postcentral gyrus</td>
<td>93</td>
<td>.011</td>
<td>4.98</td>
<td>−21</td>
<td>−37</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.80</td>
<td>−30</td>
<td>−19</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.60</td>
<td>−24</td>
<td>−31</td>
<td>52</td>
</tr>
<tr>
<td>L cerebellum</td>
<td>16</td>
<td>.679</td>
<td>4.34</td>
<td>−24</td>
<td>−55</td>
<td>−38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.04</td>
<td>−21</td>
<td>−55</td>
<td>−29</td>
</tr>
</tbody>
</table>

Regions surviving a voxel-level threshold of $p < .001$, and 10 voxels are reported for the (A) general compatibility effect and (B) Sex × Compatibility interaction for the imitation inhibition task. Subclusters at least 8 mm from the main peak are listed. **Bold** font indicates clusters that survive correction for multiple corrections using a family-wise error (FWE) correction ($p < .05$). MNI = Montreal Neurological Institute; SPL = superior parietal lobule; IPL = inferior parietal lobule; PrecG = precentral gyrus; MFG = middle frontal gyrus; L = left hemisphere; R = right hemisphere.
were asked to remember the sequence in which either four (easy condition) or eight (hard condition) digit sequences were presented on screen (see Figure 2B). After each trial, participants had to choose between two digit sequences presented numerically, one of which matched the sequence in which the digits were presented as words. Feedback was provided as to whether they answered correctly or incorrectly. The hard > easy contrast has been found to robustly activate regions of the MD network (Fedorenko et al., 2011, 2013). Each run consisted of 10 experimental blocks (each 34 sec long) and 6 fixation blocks (each 16 sec long). The total run lasted for 436 sec. Each participant completed two runs of the WM task.

The ToM localizer. To localize brain regions involved in mental state reasoning, we used a paradigm developed by Dodell-Feder and colleagues (2011; saxelab.mit.edu/superloc.php). This localizer task (see Figure 2C) includes 20 stories, each describing a false representation. Ten stories included out-of-date beliefs (the false belief condition), and the other 10 included out-of-date physical representations (photographs/maps; the false photograph condition). The false belief > false photograph contrast has been shown in prior work to robustly activate regions involved in mentalizing (Dufour et al., 2013). All trials consisted of a story (10 sec), followed by a true or false question (4 sec). Each story was separated by a 12-sec rest period. The order of the stories and conditions was the same for all participants. Each participant completed two runs of this task, with five trials per condition presented in each run.

**Behavioral Data Analysis**

RT and accuracy were recorded in the same way as Experiment 1. Compatibility effects were calculated as follows: spatial compatibility = spatially incongruent trials − spatially congruent trials; imitative compatibility = imitatively incongruent trials − imitatively congruent trials. Behavioral data were analyzed in the same fashion as Experiment 1, only separately for imitative and spatial compatibility effects. The main aim of the experiment was to test for the presence of imitative and spatial compatibility effects,
as well as for differences between the sexes (female > male). Hence, we used a one-sample $t$ test to verify the presence of spatial and imitative compatibility effects and a one-tailed independent samples $t$ test to test whether female individuals showed a higher spatial/imitative compatibility effect than male individuals.

fMRI Data Analysis

Data acquisition. Data acquisition procedures were the same as Experiment 1. There were 249 volumes collected for the imitation task, 219 for the MD network localizer, and 136 for the ToM localizer for each run.

Data preprocessing and general linear model. All MRI data were preprocessed in SPM-8. Data were realigned, unwarped, and corrected for slice timing. Data were normalized to the MNI template with a resolution of 3 mm$^3$. Normalizing to a common space instead of the individual’s native anatomical space allows for comparisons with previous studies (relying on the common space) and is preferred when definition of fROIs is based on group-constrained functional data (Nieto-Castañón & Fedorenko, 2012). Images were spatially smoothed (8 mm).

For the imitation task, a design matrix was fit for each participant with five regressors: one each for the correct trials of the four conditions and one for “new” trials (i.e., the first trial of each block). Stimulus onsets were time-locked to the presentation of the number cue with a duration of 0 sec and convolved with the standard hemodynamic response function. Contrast images were calculated for each individual participant to identify regions of the brain showing a spatial (spatially incompatible > spatially compatible) or imitative (imitatively incompatible > imitatively compatible) compatibility effect.

For the localizer tasks, the design matrix consisted of regressors for each experimental condition (“Belief” and “Photo” for the ToM localizer and “Hard” and “Easy” for the MD localizer). The onset and duration of each condition was specified and convolved with the standard hemodynamic response function. Contrast images were then calculated for each individual subject to identify regions that responded to cognitive demand (hard > easy) and mentalizing (belief > photo).

Definition of group-constrained subject-specific analyses. For the group-constrained subject-specific (GSS) analyses, the spm_ss toolbox was used, which runs in SPM using MATLAB (web.mit.edu/evelina9/www/funcloc.html). The GSS approach developed by Fedorenko et al. (2010) and Julian, Fedorenko, Webster, and Kanwisher (2012) was used to define fROIs for each participant. These fROIs were defined using (1) each individual’s activation map for the localizer tasks and (2) group-constraints or masks. These masks refer to a set of “parcels,” which demarcate areas in the brain where prior work has been shown to exhibit activity for the localizer contrasts.

Two sets of fROIs were defined (Figure 3): MD network fROIs that have been known to exhibit activity for a variety of cognitive control tasks (Fedorenko et al., 2013; Duncan, 2010) and ToM network fROIs that support mentalizing and have been specifically implicated for imitation inhibition (Brass et al., 2009; Saxe & Kanwisher, 2005). For the ToM network, four parcels were derived from a group-level map from 462 participants for the false belief > false photograph contrast (Dufour et al., 2013). These regions included the dorsal, middle, and ventral mPFC (DMPFC, MMPFC, and VMPFC, respectively) and the rTPJ. For the MD network, we used 16 parcels derived from a set of functional parcels created by Idan Blank based on a probabilistic overlap map from 197 participants (available at https://evlab.mit.edu/funcloc/download-parcels). These included areas in bilateral superior and inferior parietal lobules (SPL and IPL, respectively), inferior parietal sulcus (IPS), inferior and middle frontal gyri (IFG, MFG), precentral gyrus (PrecG),

Figure 3. Graphical representation of the parcels used to define the MD and ToM network fROIs. The MD network consisted in 16 parcels, and the ToM network included 4 parcels.
insula, and the SMA. These areas were chosen for two reasons: (1) they were part of the MD network (Fedorenko et al., 2013) and (2) they have been shown to respond in prior work to the specific type of interference control of relevance to the current study (Marsh et al., 2016; see also Experiment 1).

For each individual, these masks were used to constrain the selection of subject-specific fROIs. For each individual, for the ToM network mask, the belief > photo contrast was used, and the top 10% of voxels (based on \( t \) values) within each parcel were defined as that individual’s fROI. Similarly, for the MD network mask, each individual’s top 10% of voxels (based on \( t \) values) in the hard > easy contrast were defined as that individual’s fROI. Using the top 10% of voxels, rather than a fixed threshold (e.g., all voxels with \( p < .001 \)), ensures a constant size of each fROI across individuals (Blank, Kanwisher, & Fedorenko, 2014). We also ran the analyses using a fixed threshold (\( p < .001 \), uncorrected) and found the same pattern of results (see Supplementary Tables S1.1 and S1.2). All analyses reported below are based on the top 10% of voxels that were generated from the localizer data fROIs. Percent signal change values were extracted from all fROIs. For the main analysis, all runs of the localizer tasks were used to define fROIs in each individual. Responses in these fROIs were estimated for spatial and imitative compatibility effects.

In a supplementary analysis, responses to the localizer contrasts were also estimated to ensure that all the fROIs showed the expected response with respect to the localizer contrasts. This supplementary analysis ensured that the ToM network showed a robust belief > photo and the MD network showed a robust hard > easy effect. For these localizer analyses, an across-run cross-validation approach was used (Nieto-Castañón & Fedorenko, 2012) to ensure that data used for defining fROIs were independent of data used for estimating response (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009).

As implemented in GSS, statistical tests were performed on the percent signal change values using standard Student’s \( t \) tests. One-sample \( t \) tests were performed to investigate the response of the MD and ToM network fROIs to spatial and imitative compatibility effects. Based on prior behavioral findings, which showed greater RT interference for female than male individuals during imitation inhibition (Butler et al., 2015), we expected to observe sex differences in those regions that also show simple compatibility effects. That is, we expected brain regions that were generally involved in spatial and/or imitative control to show sex differences. As such, we only investigated sex difference in those fROIs that showed spatial or imitative compatibility effects. To do so, one-tailed independent samples \( t \) tests were performed that tested for greater engagement for female than male individuals. False discovery rate (FDR) multiple-comparison correction (\( p < .05 \)) was used to correct for the number of fROIs in each functional network.

**RESULTS**

**Experiment 1**

*Behavioral Results*

A one-sample \( t \) test confirmed a general compatibility effect (mean = 80.02, \( SE = 8.19 \), \( t(27) = 9.77, p \leq .001 \), 95% CI (63.22, 96.82), Cohen’s \( d_z = 1.85 \). A one-tailed independent samples \( t \) test showed no differences between male individuals (mean = 70.94, \( SE = 13.30 \)) and female individuals (mean = 89.10, \( SE = 9.43 \), \( t(26) = 1.114, p = .138 \), 95% mean difference = 18.16, CI (−9.64, Cohen’s \( d = .42 \.). All participants had >80% accuracy; hence, all were included in the analysis. Trials on which participants made an incorrect response (0.95%) did not make a response or responded after 2000 msec (0.52%) or pressed an invalid key or responded too fast (0.09%) were excluded from the analyses.

*fMRI Results*

In a whole-brain analysis, compatibility effects (general incompatible > general compatible) were observed in dorsomedial frontal cortex and bilaterally in dorsolateral frontal and parietal cortices (Figure 4A; Table 2A). A small volume correction (SVC) using MD and ToM network parcels was performed to restrict the search area to ToM and MD networks. Using the MD network SVC, results showed widespread activation of frontal and parietal regions, which survived correction for multiple comparisons (Figure 4A, C1). In contrast, using the ToM network SVC, no clusters survived correction for multiple comparison, and only rTPJ showed a compatibility effect at more lenient threshold (\( p < .001 \), uncorrected; see Supplementary Tables S2.1 and S2.2). Anterior mPFC did not show the general compatibility effect even at this more lenient threshold.

The Sex × Compatibility interaction revealed clusters in left SPL, extending into postcentral gyrus and a further cluster in the cerebellum (Figure 4B; Table 2B). No clusters emerged following an SVC analysis using the MD and ToM network masks, which demonstrate that the clusters emerging from the Sex × Compatibility interaction do not overlap with the MD or ToM networks (see Supplementary Tables S2.1 and S2.2; Figure 4Cii and Dii).

*Power Analysis*

We set up Experiment 1 to estimate the appropriate sample size for our critical experiment (Experiment 2). To this end, a power analysis was performed using the fMRIpower software package (fMRIpower.org; Mumford & Nichols, 2008). We performed the power analysis as follows: First, a whole-brain map of the imitation task general compatibility effect (incompatible > compatible) from Experiment 1 was entered into fMRIpower. Next, two ROIs were identified: the MD network (Duncan, 2010) and the ToM network (Saxe & Kanwisher, 2011).
The MD and ToM network masks used were the same as in Experiment 2 (see Methods). As recommended, we corrected the alpha value by the number of ROIs (0.05/2 = 0.025) before performing power analyses (Mumford, 2012).

Results from these power analyses showed that testing 50 participants in Experiment 2 would provide 80% power to detect effects as large as (or larger than) the average effect size that was observed across all nodes in the MD network in Experiment 1 (Cohen’s $d = 0.4$, mean signal change = 0.23, $SD = 0.58$). We did not have the same level of power to detect smaller effects than these, such as those observed in the ToM network in Experiment 1. Indeed, the effects in the ToM network in Experiment 1 were so small that we would have needed an impractically large sample size to achieve 80% power. As such, in Experiment 2, we decided to test participants until we had 50 usable data sets.

Design differences between Experiments 1 and 2 are worth considering when interpreting these power calculations because we may be underestimating the power of our design in Experiment 2. The toolbox used to run power calculations (fmripower.org) can only estimate power for a future experiment with the same design as the current data set (Mumford & Nichols, 2008). However, the designs of Experiments 1 and 2 differed in
two ways. First, Experiment 1 measured a general compatibility effect, whereas in Experiment 2, we broke this effect down into spatial and imitative compatibility effects. Second, Experiment 2 had more than double the amount of trials per condition as Experiment 1. Therefore, the primary contrast used to determine power was not identical to the contrast used in Experiment 2, but due to a greater number of trials per condition to estimate the effects of interest, we may underestimate power in Experiment 2. Given the lack of sex differences in Experiment 1 in our ROIs, we did not have sufficient power to convincingly investigate neural differences between male and female individuals in Experiment 2. However, given our a priori predictions regarding sex, we continue to report sex difference analyses throughout the article.

**Experiment 2**

**Behavioral Results**

The hand stimulus used in Experiment 2 for the imitation inhibition task was perceived as “neutral” by most participants (mean rating = 5.20, SD rating = 2.04; rated on a scale of 1–9, where 1 = most masculine, 5 = neutral, and 9 = most feminine). To ensure participants were engaging correctly with the task, runs on which participants had less than 80% accuracy (two runs of one participant) were removed. In addition, RTs more than 3 SDs away from the mean (two runs of one participant and one run of another participant) were excluded from the analyses. Furthermore, trials on which participants made an incorrect response (1.52%), did not make a response or responded after 2000 msec (0.61%), or pressed an invalid key (0.22%) were also excluded from the analyses. Figure 5 shows the imitative and spatial compatibility effects for both the sexes. For RT data, see Supplementary Table S3.

**Spatial compatibility.** A one-sample t test confirmed a spatial compatibility effect (mean = 41.94, SE = 2.87), t(54) = 14.618, p ≤ .001, 95% CI (36.19, 47.69), Cohen’s d$_z$ = 1.97. A one-tailed independent samples t test evidenced a greater spatial interference effect for female individuals (mean = 50.98, SE = 3.67) as compared with male individuals (mean = 33.20, SE = 3.75), t(53) = 3.38, p < .001, mean difference = 17.76, 95% CI (8.91); Cohen’s d = 0.91.

**Imitative compatibility.** A one-sample t test showed a significant imitative compatibility effect (mean = 15.37, SE = 2.86), t(54) = 5.37, p < .001, 95% CI (9.63, 21.11), Cohen’s d$_z$ = 0.72. There was no significant difference

<table>
<thead>
<tr>
<th>ROI</th>
<th>ROI Size</th>
<th>Intersubject Overlap</th>
<th>Average ROI Mask Size (Voxels)</th>
<th>Spatial Compatibility</th>
<th>Imitative Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>L_SPL</td>
<td>1173</td>
<td>1</td>
<td>117</td>
<td>2.00</td>
<td>.026</td>
</tr>
<tr>
<td>L_IPS</td>
<td>287</td>
<td>1</td>
<td>28</td>
<td>2.00</td>
<td>.026</td>
</tr>
<tr>
<td>L_IPL</td>
<td>641</td>
<td>1</td>
<td>64</td>
<td>2.72</td>
<td>.005</td>
</tr>
<tr>
<td>L_MFG</td>
<td>536</td>
<td>1</td>
<td>53</td>
<td>2.16</td>
<td>.018</td>
</tr>
<tr>
<td>L_PrecG</td>
<td>338</td>
<td>1</td>
<td>33</td>
<td>2.17</td>
<td>.018</td>
</tr>
<tr>
<td>L_IFG</td>
<td>181</td>
<td>1</td>
<td>18</td>
<td>1.83</td>
<td>.040</td>
</tr>
<tr>
<td>L_Insula</td>
<td>197</td>
<td>1</td>
<td>19</td>
<td>2.78</td>
<td>.004</td>
</tr>
<tr>
<td>L_SMA</td>
<td>294</td>
<td>1</td>
<td>29</td>
<td>2.52</td>
<td>.008</td>
</tr>
<tr>
<td>R_SPL</td>
<td>1181</td>
<td>1</td>
<td>118</td>
<td>2.30</td>
<td>.013</td>
</tr>
<tr>
<td>R_IPS</td>
<td>227</td>
<td>1</td>
<td>22</td>
<td>2.03</td>
<td>.024</td>
</tr>
<tr>
<td>R_IPL</td>
<td>599</td>
<td>1</td>
<td>59</td>
<td>2.65</td>
<td>.005</td>
</tr>
<tr>
<td>R_MFG</td>
<td>535</td>
<td>1</td>
<td>53</td>
<td>3.57</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>R_PrecG</td>
<td>269</td>
<td>1</td>
<td>26</td>
<td>2.43</td>
<td>.009</td>
</tr>
<tr>
<td>R_IFG</td>
<td>265</td>
<td>1</td>
<td>26</td>
<td>2.61</td>
<td>.006</td>
</tr>
<tr>
<td>R_Insula</td>
<td>184</td>
<td>1</td>
<td>18</td>
<td>2.09</td>
<td>.021</td>
</tr>
<tr>
<td>R_SMA</td>
<td>328</td>
<td>1</td>
<td>32</td>
<td>2.30</td>
<td>.022</td>
</tr>
</tbody>
</table>

For each individual, for the MD network mask, the hard > easy contrast was used, and the top 10% of voxels (based on t values) within each parcel were defined as that individual’s fROI. Uncorrected p values as well as FDR-corrected p values are reported. Cells in **bold** are ROIs that survive correction for multiple comparisons (p < .05, FDR-corrected).

---

Table 3. Responses in Each MD Network fROI for Spatial and Imitative Compatibility
between male individuals (mean = 15.62, \( SE = 4.39 \)) and female individuals (mean = 15.11, \( SE = 3.73 \)), \( t(53) = 0.09, p = .465 \), mean difference = -0.51, 95% CI (-10.18), Cohen’s \( d = 0.02 \).

**fMRI Results**

Five participants were excluded from the fMRI analyses due to lower than 80% accuracy in two runs of the imitation task and the MD network localizer task (\( n = 1 \)) and excessive head motion (\( n = 4 \); displacement > 4 mm) in all runs of the imitation task and/or all runs of either of the localizer tasks. Thus, the final sample consisted of 50 participants (\( M_{\text{age}} = 22.26, SD_{\text{age}} = 3.71; 24 \) female). From these 50 participants, two sessions of the imitation task were also excluded for one participant due to excessive head motion and one participant’s data for one session of the imitation task could not be used because the data file was corrupted.

**Localizer Tasks**

All fROIs showed the predicted responses to the localizer contrasts (as estimated using data not used for defining ROIs; see Methods). All the MD network fROIs showed a robust hard > easy effect (\( ts > 9.13, p < .0001 \)), and ToM network fROIs showed a robust belief > photo effect (\( ts > 5.70, p < .0001 \)). For responses for each individual fROI separately, see Supplementary Tables S4.1 (MD) and S4.2 (ToM).

**MD network fROIs**

**Spatial compatibility:** All 16 fROIs of the MD network showed a spatial compatibility effect (\( ts > 1.8, p < .04 \); Figure 6A, Table 3), which survived correction for multiple comparisons (\( p < .05 \), FDR-corrected). The mean percent signal change across the MD network for spatial compatibility was 0.70, \( SD = 1.66 \), Cohen’s \( d = 0.42 \). No significant differences were found between male and female individuals in percent signal change values in any of the fROIs (\( ts < 1.6, p > .1 \)), except right SPL which approached significance (\( p = .062 \); Figure 7A).

**Imitative compatibility:** None of the 16 MD network fROIs showed an imitative compatibility effect, which...
survived correction for multiple comparisons (all \(p > .05\), FDR-corrected). Five MD network fROIs showed an imitative compatibility effect at an uncorrected threshold \((t > 1.95, p < .05)\). These fROIs include bilateral IPL, bilateral IPS, and the right IFG (Figure 6A, Table 3). Four further fROIs showed an imitative compatibility effect that approached significance, which included left IFG \((p = .07)\), right SPL, right MFG, and right PrecG \((p = .06)\). The mean percent signal change across the MD network for imitative compatibility was 0.54, SD = 2.06, Cohen’s \(d = 0.26\). There was no significant difference between male and female individuals in any of these fROIs \((t < 1.5, p > .08);\) see Figure 7B).

**ToM network fROIs.** None of the ToM network fROIs showed imitative \((t < 1.3, p > .50)\) or spatial \((t < 1.6, p > .06)\) compatibility effects, even at an uncorrected significance threshold (Figure 6B, Table 4). rTPJ showed a spatial compatibility effect that approached significance \((p = .065)\). The mean percent signal change across the ToM network for spatial compatibility was \(-0.16, SD = 1.88\), Cohen’s \(d = -0.08\), and the mean percent signal change across the ToM network for imitative compatibility was \(-0.32, SD = 2.02\), Cohen’s \(d = -0.16\).

**Whole-brain analyses.** For completeness and for use in future meta-analyses, we also computed group-level whole-brain analyses separately for general, spatial, and imitative compatibility effects, as well as for Sex \(\times\) Compatibility interactions (see Supplementary Table S4).

**Open Science**

Data for Experiments 1 and 2 are freely available online including behavioral and fROI data (osf.io/45x6z), as well as whole-brain \(t\) maps (https://neurovault.org/collections/3218).

**Table 4.** Responses in Each ToM Network ROI for Spatial and Imitative Compatibility

<table>
<thead>
<tr>
<th>ROI</th>
<th>ROI Size</th>
<th>Intersubject Overlap</th>
<th>Average ROI Mask Size (Voxels)</th>
<th>Spatial Compatibility</th>
<th>Imitative Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(t) (p) (p)-FDR (t) (p) (p)-FDR</td>
<td></td>
</tr>
<tr>
<td>DMPFC</td>
<td>576</td>
<td>1</td>
<td>57</td>
<td>-1.38 (.913) (.951)</td>
<td>-1.167 (.876) (.898)</td>
</tr>
<tr>
<td>MMPFC</td>
<td>494</td>
<td>1</td>
<td>49</td>
<td>-0.043 (.517) (.951)</td>
<td>-1.081 (.857) (.898)</td>
</tr>
<tr>
<td>VMPFC</td>
<td>382</td>
<td>1</td>
<td>38</td>
<td>-1.690 (.951) (.951)</td>
<td>-1.286 (.898) (.898)</td>
</tr>
<tr>
<td>rTPJ</td>
<td>1018</td>
<td>1</td>
<td>101</td>
<td>1.543 (.665) (.258)</td>
<td>-0.106 (.542) (.898)</td>
</tr>
</tbody>
</table>

For each individual, for the ToM network mask, the Belief > Photo contrast was used, and the top 10% of voxels (based on \(t\) values) within each parcel were defined as that individual’s fROI. Uncorrected \(p\) values as well as FDR-corrected \(p\) values are reported.
DISCUSSION

The current study provides the most robust neuro-imaging evidence to date for a lack of functional specificity in the neural circuits supporting the inhibition of automatic imitation. With higher statistical power and functional sensitivity than prior studies, across two experiments the results demonstrate that imitation inhibition engages a domain-general neural network as opposed to a brain network that supports social cognition. As such, models of imitation control need updating to include an increased role for domain-general processes and a reduced or altered role for domain-specific processes. Furthermore, in terms of behavior, female individuals showed a higher spatial but not imitative compatibility effect than male individuals. However, there was no sex difference in the neural mechanisms underlying spatial or imitation control, which suggests that further exploration of sex differences in inhibitory control is required.

Functional Specificity in Imitation Inhibition

Our findings show that brain regions that are engaged in a verbal WM task, which are associated with the operation of the MD network (Fedorenko et al., 2013; Duncan, 2010), are also engaged during spatial and imitative conflict resolution. These results support the involvement of a domain-general cognitive and neural system during the control of imitation. By contrast, brain regions that are engaged in a belief reasoning task, which are associated with the operation of the ToM network (Van Overwalle, 2009; Saxe & Kanwisher, 2003; Frith & Frith, 1999), show no engagement during the inhibition of imitation. As such, we provide no evidence for domain specificity in cognitive and neural systems that control imitation.

Brass and colleagues (2009) proposed that, in the context of imitation control, rTPJ is involved in self–other distinction, and mPFC enforces the self-generated action over the observed action. Our findings are inconsistent with the hypothesis that a specific neural system related to social cognition is engaged in the inhibition of automatic imitative tendencies. mPFC and rTPJ have both been implicated in imitation inhibition by some studies (Wang et al., 2011; Brass et al., 2009; Spengler et al., 2009; Brass, Derrfuss, & von Cramon, 2005). In contrast, other studies found engagement of mPFC only (Cross et al., 2013; Brass et al., 2001) or of domain-general regions rather than mPFC and rTPJ (Marsh et al., 2016; Cross & Iacoboni, 2013; Crescentini et al., 2011; Bien et al., 2009). In both experiments in the current study, we had larger sample sizes than prior experiments, and in Experiment 2, we had sufficient statistical power to be confident in detecting effects as large as previously observed in mPFC and rTPJ, should they exist. Taken together with prior findings (Table 1), we suggest that, during the inhibition of imitation, the consistency of mPFC and rTPJ engagement across individuals is relatively low, whereas the consistency of MD network engagement across individuals is relatively high.

These results have potential implications for self–other control theories of social cognition more generally. Mostly based on imitation research, which previously suggested that mPFC and rTPJ are engaged in imitation inhibition, self–other control is thought to be a candidate mechanism for a diverse set of social functions (de Guzman et al., 2016; Sowden & Shah, 2014; Brass et al., 2009). For example, self–other control processes have been linked to autism, empathy, and theory of mind (de Guzman et al., 2016; Sowden & Shah, 2014; Spengler et al., 2009). However, recent behavioral findings, which used larger sample sizes than prior work and meta-analytical approaches, do not support the view that the control of imitation varies as a function of social disposition as indexed by autistic-like traits and empathy (Cracco et al., 2018; Genschow et al., 2017; Butler et al., 2015). In light of these recent behavioral results, the lack of engagement of mPFC and rTPJ in the current study raises an important question about the reliance of imitation inhibition on a self–other distinction. One possibility is that, instead of a distinctly social mechanism (Bertenthal & Scheutz, 2013; Boyer et al., 2012), inhibiting imitative tendencies may involve the same cognitive processes that are used when inhibiting other nonsocial external influences (Cooper et al., 2012; Heyes, 2011).

Alternatively, the engagement of mPFC and rTPJ during self–other control processes may be more complicated than current models of social cognition suggest. Indeed, a small number of neurostimulation studies have shown that modulation to rTPJ can influence performance on RT measures of imitation (Sowden & Catmur, 2015; Hogeveen et al., 2014). In addition, mPFC and rTPJ have been found to be involved in the modulation of automatic imitation. For example, Klapper et al. (2014) found a higher response in rTPJ when an interaction partner looked human and was believed to be human compared with when neither of these animacy cues was present. Wang and colleagues (2011) demonstrated that mPFC had a top–down influence on other brain circuits during social modulation of imitation via direct gaze. These studies suggest that mPFC and/or rTPJ may have a regulatory role, be sensitive to social context, and be functionally connected to other regions during the inhibition of automatic imitation. Indeed, regions that do not show direct engagement in a cognitive process of interest have been known to have a regulatory influence on other regions that are directly engaged (Burnett & Blakemore, 2009). In line with this proposal, Cross and colleagues (2013) suggested that imitation control involves top–down regulation between a domain-general cognitive control network and a domain-specific network relevant for imitation. More generally, research from other domains of social cognition shows growing evidence for higher complexity and functional interplay within and between so-called domain-specific...
and domain-general networks (Spunt & Adolphs, 2015; Baetens, Ma, Steen, & Van Overwalle, 2014; Quadflieg et al., 2011; Zaki, Hennigan, Weber, & Ochsner, 2010). These studies suggest that models including next divisions between these networks may be an overly simplistic characterization of mental function (Michael & D’Ausilio, 2015; Barrett, 2012). Much like social cognition in general, therefore, imitation control may be best explained by interactions between component functional circuits, which themselves need not be domain-specific (Spunt & Adolphs, 2017). A crucial direction for future research is testing for more complex models of imitation, which may involve connectivity in and between regions of the MD and ToM networks.

An important point to note, however, is that any conclusions made regarding possible domain specificity of mPFC and rTPJ are based on the assumption that mPFC and rTPJ are at least partly specialized for social cognition (Brass et al., 2009). Recent evidence suggests that mPFC and rTPJ may be functionally versatile in the sense that they show general cognitive properties, which may not be specific to social cognition (Dugué, Merriam, Heeger, & Carrasco, 2017; Schurz, Tholen, Perner, Mars, Sallet, 2017; Schuwerk, Schurz, Müller, Rupprecht, & Sommer, 2017; de la Vega, Chang, Banich, Wager, & Yarkoni, 2016; Carter & Huettel, 2013; Alexander & Brown, 2011; Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). Thus, the argument that the engagement of mPFC and rTPJ in imitation inhibition may be specific to social cognition might need further validation. In addition, social cognition itself has been broken down in “bottom–up” and “top–down domains (Zaki & Ochsner, 2012). The bottom–up domain refers to prereflective processes that are fast and stimulus driven, whereas the top–down domain maps on reflective, cognitively laboratory, and flexible processes (Bohl & van den Bos, 2012). When extended to imitation control, prior research has consistently implicated regions involved in top–down control for automatic imitation (Brass et al., 2009). However, recent studies suggest that imitation control (and social cognition more broadly) relies on interactions between bottom–up and top–down processes (Christov-Moore, Conway, & Iacoboni, 2017; Cross & Iacoboni, 2014; Bohl & van den Bos, 2012). Thus, another important avenue for future research would be to investigate imitation control based on bottom–up and top–down processes and their interactions, rather than considering these processes as mutually exclusive.

Nonetheless, results from the current study remain clear: The basic imitation inhibition mechanism engages the MD network, which has been consistently associated with domain-general processes (Duncan, 2010). Given the mixed findings in prior imitation studies (Table 1) as well as in psychology and neuroscience more generally (Open Science Framework, 2015; Button et al., 2013), future fMRI research may also consider reliability and reproducibility as key concerns in imitation research and consider the possible use of fROI approaches as a means to quantify consistency across individuals.

Sex Differences in Imitation Inhibition

This study is the first to investigate sex differences in the neural mechanisms that inhibit imitation. The behavioral data demonstrated that female individuals show a greater spatial but not imitative compatibility effect than male individuals. This result extends prior behavioral research on sex differences, which did not separate spatial (or orthogonal spatial) from imitative responses in imitation control (Gensche et al., 2017; Butler et al., 2015). The result is also consistent with reports in a wide range of non-social inhibitory control tasks, which show similar sex differences (Figure 2; Stoet, 2010, 2017; Clayson et al., 2011; Rubia et al., 2010; Bayliss et al., 2005). All these tasks share a common feature—they require the inhibition of a response to a task-irrelevant spatial feature to enforce a task-relevant response. Taken together, this pattern of results suggests that response inhibition relating to spatial conflict differs between the sexes, rather than a process that is tied to the control of imitation. An alternative possibility is that the difference between the sexes for spatial compatibility is larger than for imitative compatibility, and we were unable to detect the imitative effect behaviorally. Future research will have to probe these possibilities further.

Given the proposed role of MD and ToM networks in imitation control, we anticipated sex differences in one or both of these networks. The neuroimaging data, however, demonstrated no sex differences in the ToM or MD networks in either experiment. Furthermore, even though regions outside our ROIs mediated the sex difference in Experiment 1, these regions were not consistently engaged differently for male and female individuals in Experiment 2. Thus, based on data across both experiments, our best estimate is that univariate analyses, which assess the magnitude of BOLD response, do not show large effects of sex in MD or ToM neural networks. This being said, there does seem to be a trend for greater engagement in the MD network for female individuals compared with male individuals for both spatial and imitative effects, but this does not survive our statistical thresholding (Figure 7). As a consequence, we are cautious to interpret this null result as we did not have the same level of statistical power to detect sex differences as we did to detect simple compatibility effects. Indeed, it remains a possibility that small univariate effects exist or that the sex difference is underpinned by more complex neural organization. Future studies that use connectivity measures (Spoorns, Tononi, & Köttler, 2005) or multivoxel pattern analysis (Kriegeskorte, Mur, & Bandettini, 2008; Norman, Polyn, Detre, & Haxby, 2006) may show increased sensitivity and be better able to capture the complexity of neural organization that we are aiming to measure.

Limitations

The primary limitation of the current work is that we studied a relatively simple model of brain organization...
based on univariate measures. Given the mixed evidence from prior studies regarding imitation control (Table 1), we felt it was an important step to first establish the extent to which general and specific systems were engaged in a univariate manner. By doing so, we aimed to build an appropriate foundation for future work to build upon. Moreover, as we only identified the MD and ToM networks, it is possible that neural regions outside our key networks may play a role in imitation inhibition or mediate the sex difference in spatial response inhibition or imitation control. Even though our whole-brain analyses showed no consistent effects outside our fROIs, this only shows that there was no univariate engagement of extended brain regions. We thus acknowledge that we have tested a relatively simple model of brain organization that is likely to underestimate the complexity of neural processes associated with social and cognitive mechanisms such as imitation control. As mentioned before, future work may consider interactions between general and specific systems and more complex, multivariate measures of brain organization.

A second limitation regards the functional localization approach used to identify the ToM and MD networks in Experiment 2. The validity of the fROI approach is based on assumptions about the functional processes that are engaged by the localizers used to identify fROIs. For example, different ToM localizers may engage partly nonoverlapping aspects of the ToM network (Schaafsma, Pfaff, Spunt, & Adolphs, 2015; Spunt & Adolphs, 2014). Therefore, our conclusions about the role of ToM and MD networks are limited to the type of localizer paradigms that we used in the current study. Future research that uses different functional partitions of these networks would be instructive.

A final potential limitation is that the order of tasks in Experiment 2 could have influenced our results. We ordered the tasks such that the ToM localizer was always performed at the end, but the MD task was interspersed between imitations runs in order to offset boredom. We arranged blocks in this manner because we were primarily concerned that asking people to perform a belief reasoning task would introduce a social bias to treat the person (hand image) in an artificially more social/belief reasoning manner during the imitation inhibition task. We did not share the same level of concern that performance in the current experiment that the MD task influenced the way the imitation task was performed. This being said, we did get the same results in Experiment 1, when the MD task was not performed before the imitation task. As such, although possible, we find it unlikely that task order had a meaningful impact on our results in Experiment 2.

Acknowledgments

This work was funded by a grant from the Economic and Social Research Council (Grant ES/K001884/1 to R. R.). We thank Ruud Hortensius for comments on a previous version of the manuscript.

Reprint requests should be sent to Richard Ramsey, Department of Psychology, Bangor University College of Health and Behavioural Sciences, Adeilad Brigantia, Penrallt Road, Bangor, Gwynedd, LL57 2AS, United Kingdom, or via e-mail: r.ramsey@bangor.ac.uk.

REFERENCES


Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrising theory of mind and interactionism to study the interplay of Type 1 and Type 2 processes. *Frontiers in Human Neuroscience*, 6, 274.


