# SCIENTIFIC REPORTS

Received: 24 August 2016 Accepted: 28 March 2017 Published online: 28 April 2017

## **OPEN** Neural Activity while Imitating **Emotional Faces is Related to Both** Lower and Higher-Level Social **Cognitive Performance**

Colin Hawco<sup>1</sup>, Natasa Kovacevic<sup>2</sup>, Anil K. Malhotra<sup>3</sup>, Robert W. Buchanan<sup>4</sup>, Joseph D. Viviano<sup>1</sup>, Marco Iacoboni<sup>5</sup>, Anthony R. McIntosh<sup>2</sup> & Aristotle N. Voineskos<sup>1</sup>

Imitation and observation of actions and facial emotional expressions activates the human fronto-parietal mirror network. There is skepticism regarding the role of this low-level network in more complex high-level social behaviour. We sought to test whether neural activation during an observation/imitation task was related to both lower and higher level social cognition. We employed an established observe/imitate task of emotional faces during functional MRI in 28 healthy adults, with final analyses based on 20 individuals following extensive guality control. Partial least squares (PLS) identified patterns of relationships between spatial activation and a battery of objective outof-scanner assessments that index lower and higher-level social cognitive performance, including the Penn emotion recognition task, reading the mind in the eyes, the awareness of social inference test (TASIT) parts 1, 2, and 3, and the relationships across domains (RAD) test. Strikingly, activity in limbic, right inferior frontal, and inferior parietal areas during imitation of emotional faces correlated with performance on emotion evaluation (TASIT1), social inference - minimal (TASIT2), social inference enriched (TASIT3), and the RAD tests. These results show a role for this network in both lower-level and higher-level social cognitive processes which are collectively critical for social functioning in everyday life.

Social interactions are complex processes which are built upon the perception and understanding of the actions, intentions, and mental state of others. Social cognitive neuroscience has suggested a dichotomy between low-level processes such as understanding motor actions or emotional processing and higher-level processes such as inferring the mental states of others (e.g. theory of mind) or detecting deception and sarcasm. This apparent dichotomy between lower level and higher level social cognitive processes is supported by behavioral<sup>1</sup>, neuroimaging<sup>2</sup> and lesion studies<sup>3,4</sup>. Neuroimaging studies have primarily examined these processes with two approaches: using tasks relevant to the perception of and interaction with other people's actions (action observation and imitation), and using tasks relevant to infer the beliefs and desires of the other person (mentalizing). Lower level social cognitive processes such as action interpretation have been associated with activation of a lateral frontal-parietal network, the putative human analogue to the mirror neuron system observed in monkeys<sup>2,5</sup>. Higher-level social cognitive processes, such as theory of mind, have been associated with activation of cortical midline regions, including the medial prefrontal cortex, posterior cingulate cortex, and precuneus, as well as temporoparietal junction and temporal pole. This network is also known as the 'mentalizing' system. Dual processing models have been proposed in which the lateral fronto-parietal network supports relatively automatic processes such as identifying motor actions while the mentalizing network supports controlled processes of attributing actions to complex social causes<sup>6-8</sup>.

<sup>1</sup>Campbell Family Mental Health Institute, Centre for Addiction and Mental Health, Department of Psychiatry, University of Toronto, 250 College St., Toronto, M5T 1R8, ON, Canada. <sup>2</sup>Rotman Research Institute, Baycrest Geriatric Centre, 3560 Bathurst Street, Toronto, M6A 2E1, Ontario, Canada. <sup>3</sup>Zucker Hillside Hospital, 75-59 263rd St, Glen Oaks, 11004, NY, United States. <sup>4</sup>Maryland Psychiatric Research Center, 55 Wade Ave, Catonsville, 21228, MD, United States. <sup>5</sup>University of California – Los Angeles, Los Angeles, 90095, CA, United States. Correspondence and requests for materials should be addressed to A.N.V. (email: Aristotle.voineskos@camh.ca)

Most neuroimaging studies examining the lateral fronto-parietal network have made use of hand actions or other forms of simple motor stimuli<sup>9</sup>. As such, the role of this system has been mainly considered in the context of perceiving gross motor actions. Some have taken the concept of mirroring one step further, by arguing that the lateral fronto-parietal network is important for emotional empathy by permitting humans to feel what others feel, and potentially to assess or interpret emotional cues and social behaviour<sup>10-12</sup>. Investigating this issue may be best served by paradigms utilizing socially relevant stimuli as opposed to simple motor acts. One such paradigm is the imitation and observation of emotional faces, which has been shown to activate the lateral fronto-parietal 'mirror' network<sup>13-15</sup>. This is consistent with the notion that activation of this network can play a role in allowing people to empathize by imitating emotions<sup>11, 12</sup>. Activity in the right IFG has been positively correlated with self-report measures of empathy in healthy children<sup>15</sup>, and has been shown to be more active in normally developing children than those with autism<sup>13</sup>, with activity in autistic children inversely correlated with social impairment. However, the IFG has also been implicated in some higher level social cognitive processes, specifically the inhibiting self-perspective in social judgements<sup>16, 17</sup>. Recent work has also shown structural changes in right fronto-parietal cortex in more socially impaired people with schizophrenia<sup>18, 19</sup>, a group of individuals in whom higher-level social cognitive processes such as theory of mind are more prominently affected<sup>20</sup>. These findings raise the question of whether neural activation during basic mirroring tasks is related to higher-level social cognitive abilities.

Meta-analysis of fMRI data has shown little overlap between activity in the brain regions associated with the mirroring and mentalizing systems across a range of tasks<sup>2</sup>, and brain lesion studies have suggested a dissociation between mirroring and mentalizing regions<sup>3,4</sup>. As such, mirroring and mentalizing were initially thought to represent dichotomous systems which function relatively independently, and the role of mirroring in higher-level social cognition is a contentious issue<sup>21–23</sup>. However, there is a growing body of evidence that the mentalizing and mirroring systems may interact during more complex social cognitive processing, such as watching social videos<sup>24–27</sup>. Co-activation of mentalizing and mirroring regions has also been noted when viewing agents performing irrational actions, suggesting aspects of the mentalizing system may facilitate interpreting unexpected actions<sup>28–30</sup>. The mirroring and mentalizing system may work together during social cognitive processing, although a specific role of the mirror system in higher-level cognitive processing has yet to be established.

The purpose of this study is to determine whether brain regions activated during a facial emotion mirroring task are related to lower-level social cognitive performance, higher-level social cognitive performance, or both. We utilized a process specific social mirroring task (the facial imitate/observe task<sup>13-15</sup>) inside the fMRI, along with a battery of objective out-of-scanner social tests ranging from basic emotional-perception tasks to complex higher level social cognitive tasks such as detection of lies or sarcasm. Brain-behavior relationships were examined using the spatio-temporal partial least squares (PLS), a non-parametric multivariate approach<sup>31, 32</sup>. PLS identifies latent variables (LVs), linear combinations of fMRI signal amplitude at each voxel across time and a design matrix which can consist of either experimental conditions or correlations across conditions with behavioral scores. Each LV expresses a pattern of the common covariance between the design and the fMRI data. PLS is designed to handle data with high collinearity while simultaneously capturing essential non-redundant relationships among the data<sup>31</sup>, overcoming the confounding influence of collinear variables in traditional multiple regression<sup>33</sup>. This makes PLS an excellent approach to uncover relationships between brain activity and a set of social cognitive scores which are likely highly interrelated. We hypothesized that a) consistent with previous studies, imitating emotional faces would activate the mirroring network, more prominently in the imitate than the observe condition, and b) neural activity during imitation of emotional faces specifically (rather than imitating neutral faces or observing faces) would be associated with both lower-level and higher-level social cognitive performance.

#### Results

Task Effects. Task PLS is a multivariate model-free approach to identify spatial patterns which share relationships across the experimental conditions. A task PLS was run on the data to replicate previous task-based analysis, and examine the underlying pattern of task task-evoked spatial activity. PLS generated 6 LVs (as there were three trial types, emotional faces, neutral faces, and fixation, and two conditions, Imitate and Observe). Permutation analysis (see methods below) showed significance (p < 0.05) in the first three LVs, which accounted for 61.6%, and 14.5%, and 12.9% of crossblock covariance respectively (Fig. 1). LV 1 was related to viewing faces as opposed to fixation trials, and showed a pattern related to the lateral frontal-parietal network similar to the results of previous studies<sup>13–15</sup>. LV 1 was related to viewing faces in general rather than distinguishing between emotional and neutral faces, demonstrating that this pattern of faces >fixation in the mirroring network is the dominant pattern of neural activity within this task analysis. LVs 2 and 3 explain some of the remaining variance in the data not accounted for the pattern presented in LV 1, representing interaction effects showing different patterns of activity across the Imitate and Observe conditions. LV 2 shows a large amount of overlap with LV 1, suggesting some of the variance within those regions is explained by aspects of both LVs 1 and 2. The majority of the variance explained by LV 1, but some additional variance within those voxels is explained by the pattern of emotional faces > neutral and fixation in Imitate and increased activity in fixation trials for Observe, as seen in LV 2. LV 3 represents an interaction in which the relative effects of neutral faces and fixation are different in the Imitate and Observe tasks.

For consistency and replication of previous studies, we also ran a general linear model (GLM) in SPM8. The GLM showed a similar pattern as in previous studies and LV1 of the task PLS (see supplementary methods and Supplementary Figure 1).

**Behavioral PLS.** Behavioral PLS examined the relationships between patterns of activity in the brain and social cognitive test scores. Scores from six social cognitive tasks were included: the Emotion Recognition



**Figure 1.** Results from the first latent variable using task PLS analysis. Data is shown for LV1, 2 and 3. The top panel shows the design pattern (the contribution of each condition to the LV). Error bars are 95% confidence interval derived from the bootstrap analysis. The bottom panel shows the voxel pattern from the first lag of the PLS analysis rendered on the cortex. Sagittal slices are shown for X = -10 (left) and X = 10 (right). Voxel intensity is displayed as bootstrap ratio, a measure of reliability of voxels within the LV. Red-yellow regions show the pattern described by the design pattern, while blue regions show the opposite pattern.

.....

(ER40)<sup>34</sup>, the Reading the Mind in the Eyes Test, RMET<sup>35</sup>, the Relationships Across Domains test (RAD)<sup>36</sup>, and the three subtests of the Awareness of Social Inference Test Revised (TASIT-R)<sup>37</sup>. TASIT1 can be considered a test of lower-level social cognitive functioning, while TASIT2 and TASIT3 are high-level tests. The behavioral PLS generated 36 LVs, as there were six experimental conditions (emotional faces, neutral faces, and fixation, separately for Imitate and Observe, and six social cognitive tests). Overall permutation analysis showed significance in the first two LVs, which accounted for 23% and 13.2% of crossblock covariance respectively. However, the split-half permutation reliability analysis (see methods) demonstrated that the relationship between the design pattern and brain pattern was not reliable within LV 1 and as such it was not considered further. In LV 2, neural activity during imitation of emotional faces was positively correlated with RAD, TASIT1, TASIT2, and TASIT3 (Fig. 2, top panel). Neural activity for Fixation and Neutral faces during Imitate, as well as Neutral faces during Observe, was negatively correlated with those same test scores. As such, LV 2 shows a pattern of correlations during the imitation of emotional faces related to performance on TASIT1, TASIT2, TASIT3, and RAD.

The pattern of brain activity in LV 2 (Fig. 2, bottom panel) when imitating emotional faces in lag 1 showed a positive relationship with performance on TASIT1 TASIT2, TASIT3, and RAD. Activity was present in right pars opercularis (within inferior frontal gyrus), right and left supramarginal gyrus, right motor cortex, bilateral anterior temporal regions, fusiform gyrus, and parahippocampal cortex. Activity was also noted in right anterior/mid cingulate, right rolandic operculum, right putamen, and the cerebellum. Activity in lags 4 and 5, which positively correlated with social cognitive performance, was notably present in anterior temporal regions as well as in posterior middle and inferior right temporal gyri. Several regions showed negative relationships to the design pattern (thus greater fMRI activity while imitating emotional faces in individuals with poorer social cognitive performance), including the left inferior frontal gyrus (pars operculum and pars triangularis), and bilaterally in the basal ganglia (greatest in left putamen and right globus pallidus).

#### Discussion

The behavioral PLS results of this study showed a robust relationship between neural activity during a process-specific imitate-observe task and both lower-level and higher-level social cognitive performance. While a growing number of studies using complex stimuli have noted co-activation of mirroring and mentalizing regions<sup>24, 27, 38</sup>, co-activity cannot be taken to conclude that mirroring has a strong role in higher level social cognition. The novelty of our study is that we showed complex relationships of neural activity during a simple and process-specific mirroring task with performance on objective out-of-scanner assessments of higher-level social cognitive performance. In addition, the PLS approach allowed us to include several assessments in a single analysis, providing data driven evidence of which assessments best capture these brain-behaviour relationships. These behavioral relationships were specific to imitation of emotional faces, rather than neutral faces. Several regions outside the right fronto-parietal circuit, including left IFG and some mentalizing regions (e.g. TPJ), also showed negative correlations with social cognitive performance, suggesting either compensatory responses or over-activation in participants with poorer social cognition.



**Figure 2.** LV 2 of the behavioral PLS results. (**A**) Design pattern showing the pattern of correlations between brain signal for each during event type and each behavioral score within the voxel pattern. Error bars represent 95% confidence interval based off a bootstrapping analysis of 1000 iterations. (**B**) Voxel patterns for each lag, displayed as bootstrap ratio on the MNI152 brain. Data is shown for two sagittal views as well as rendered onto the cortex (8 mm search depth) to better visualize cortical activity. Red-yellow shows a correlation pattern matching the design pattern (**A**), while blue regions show the opposite pattern. Each lag represents a time point (TR) following stimuli onset (which occurred during 'Lag0').

These findings of a relationship between activity in a simple mirroring task and higher level social cognitive processing scores suggest a relationship for the mirror network in higher-order social cognitive processing including complex tasks. Effective mentalizing in humans although likely primarily reliant on classic mentalizing regions may be further subserved by mirroring regions<sup>39, 40</sup>. Given that the relationships elicited were noted in assessments involving interpretation of dynamic interactive video tasks including human actors, it may be that such relationships are only evoked when we consider social cognition as a process involving integrating information from multiple processing levels. The mirroring network may not be solely for understanding the intentions and physical goals of motor acts<sup>41, 42</sup>. Our findings add to the debate regarding the role of this network in social cognition<sup>2, 21, 22, 43, 44</sup>. Several authors have argued that this network is important not only for imitation or action perception, but also for emotional empathy and possibly even in cognitive empathy (sometimes used interchangeably with theory of mind)<sup>23, 42</sup>, consistent with the theory of embodied simulation<sup>10, 44</sup>, despite debate to the contrary<sup>21, 22, 43</sup>. Through the combination of a process-specific in-scanner task, more naturalistic and objective social cognitive tests, and the PLS multivariate framework, we were able to uncover essential brain-behavior relationships for which traditional task contrasts or linear regression analyses may be less sensitive.

Higher level social cognitive processing involves concepts related to both theory of mind and cognitive empathy, both complex constructs which likely encompass several mental processes. To date, there is only limited evidence of overlap in neural activity between high-level tasks related to the mentalizing network and low-level tasks related to the mirroring network<sup>2, 21, 22, 43, 45, 46</sup>. However, task based analysis is often performed by contrasting conditions, and many mirroring tasks have utilized stimuli with little or no direct social relevance, or tasks designed to evaluate highly specific social cognitive processes which may not be well suited to capturing overlap within these systems<sup>47</sup>. By moving away from an 'overlapping activity' perspective into an approach focused on brain-behavior relationships, we were able to demonstrate that activity in the mirroring network is correlated with higher-level social cognitive abilities. This combination of pairing a process-specific task paradigm with dynamic social cognitive measures may be a powerful approach for probing relationships between social cognitive networks and real-life social cognitive abilities.

The behavioral PLS LV correlating neural activation during imitation of emotional faces with TASIT and RAD scores also included some regions which are often considered part of the mentalizing network. While these regions did not robustly activate during the task PLS, their presence in the behavioural PLS reflects activation during the imitate task that is relevant for social cognitive performance. Positive relationships with social cognition and emotional faces were observed in the temporal pole bilaterally, a region implicated in mentalizing and theory of mind<sup>48</sup> as well as in processing deception<sup>49</sup>. Activity was also noted in the posterior region of the superior temporal sulcus, which may be of particular interest as this area has been implicated in both the mirroring/ imitation network<sup>50</sup> and the mentalizing network<sup>48</sup>. The superior temporal sulcus may serve as an integrative region, sharing information between these two networks<sup>51</sup> and/or as serving as a relay point for high level visual information<sup>55, 56</sup>, and gaze perception<sup>2</sup>. As such, this region is involved with perceptual processes necessary for both low-level and high-level social cognition. It remains to be seen if this region may serve as an integrative hub between the mentalizing and mirroring networks, though the posterior superior temporal sulcus shares connectivity within both networks<sup>51</sup> and has been implicated as an important region in social cognitive impairment in autism<sup>57, 58</sup>.

Studies showing increased connectivity between the right IFG (a critical hub of the mirroring network) and the mentalizing network provide evidence for functional communications between these networks<sup>24, 59–62</sup>. This connectivity between mirroring and mentalizing regions has been shown to be disrupted in autism<sup>59, 63</sup>, raising the possibility that disrupted social cognition in autism may be driven not only by localized effects within one of these systems but also as a network level disruption in how these systems interact. Likewise, social cognitive deficits are increasingly recognized as a core feature of schizophrenia, strongly related to overall functional outcomes<sup>1</sup>. People with schizophrenia have been noted to show reduced right IFG activity when imitating emotional faces as well as opposite relationships between self-reported empathy and right IFG activity compared to controls<sup>64</sup>. As social cognitive impairment is present at, or potentially preceding the onset of psychosis<sup>20</sup>, it is possible this relatively brief imitate/observe in-scanner task may be useful as a potential early risk marker of the disease.

We found that increased activity in the left IFG was associated with poorer social cognitive performance, as well as in clusters in the left medial frontal cortex and the TPI<sup>4, 46-48</sup>. One study examining resting state network connectivity noted over-connectivity of the left IFG in the mirroring network in participants with autism, consistent with our finding that over activity in the left IFG is associated with poorer social cognitive performance<sup>65</sup>. A number of regions outside the mirroring or mentalizing networks also showed a negative relationship with social cognitive performance, including the basal ganglia, thalamus, occipital cortex and superior colliculus. These regions may be involved in emotional face processing<sup>66</sup>. We propose three possible explanations for the negative correlations between this pattern of neural activity when imitating emotional faces and social cognitive performance: 1) prolonged neural activity related to less efficient processing resulting in a greater magnitude of hemodynamic signal, as may be suggested by increased activity in regions associated with visual and face processing; 2) compensatory activity in which individuals with less efficient processing make greater use of the left IFG to compensate for lack of activity on the right (the compensatory hypothesis); or 3) over activity in the left IFG resulting in competition with right IFG, resulting in interference or an decrease in network coherence and deficits in social cognitive test performance (the dedifferentiation hypothesis). One possible explanation is that participants with poor performance may be attempting to compensate by activating aspects of the mentalizing system during lower-level social cognitive processing. It has been suggested that the anterior cingulate cortex, a region noted in the behavioral PLS, is involved in biasing activity towards either mentalizing or simulation<sup>26</sup>.

We did not find reliable relationships between the ER40 and RMET with neural activity to emotional faces. The ER40 and RMET rely on static images which fail to capture the full complexity of social cognitive processing, unlike the TASIT which requires interpretation of video scenarios of complex human interactions and the RAD which involves interpreting stories. It may be that more complex and process-general tests are required to more fully elicit relationships between activity in the simulation and mentalizing networks<sup>26, 47</sup>. Supporting this assertion, correlations between TASIT test scores and the observed voxel pattern in the PLS analysis were as high as 0.8, indicating a very strong relationship between mirroring network activity and social cognitive performance measured by TASIT1 (a lower-level task), as well as higher-level tasks TASIT2, and TASIT3. Some recent studies using more naturalistic higher-level social-cognitive paradigms, such as videos, have implicated IFG activity<sup>25, 26</sup>, and a small number of recent studies involving more complex social interaction tasks have noted co-activation in mirroring and mentalizing regions<sup>24, 27, 38</sup>.

Here we provide evidence that neural activation while imitating emotional faces is related to performance on dynamic and objective 'real-life' assessments of even the most complex social cognitive tasks. Given the relatively

small sample size, further research expanding into a larger sample would be worthwhile to replicate and extend these findings. However, our findings suggest that patterns of neural activation during basic imitative behaviour may be a surrogate marker for highly complex social function in humans. The paradigm demonstrated here may be particularly useful for early identification studies in neurologic and psychiatric disorders with social cognitive impairment, and in intervention studies using 'target engagement' as a marker of early treatment response.

### Methods

**Participants.** Twenty-eight healthy adult participants were initially included in this study. Average age of participants was  $34.3 \pm 11.7$  years, with 18 men and 10 women, 23 right-handed participants, and an average level of education of  $15.1 \pm 2.22$  years. Participants aged 18 to 55 were included in the study. All participants completed the Edinburgh handedness Inventory and were administered the Structured Clinical Interview for DSM-IV disorders (SCID) to rule out possible psychiatric illness. Urine toxicology screening was performed to further ensure that no participant with a current substance use disorder was included in the study. Additionally, exclusion criteria included a first degree relative with a history of psychotic mental disorder, a history of head trauma resulting in unconsciousness, or a history of seizure or other neurological disorders. The protocol was approved by the research ethics board of the Centre for Addiction and Mental Health, University of Toronto, all research was conducted in accordance with the declaration of Helsinki and the tri-council policy statement on Ethic Conduct for Research Involving Humans. All participants gave informed consent, and signed an institutionally approved informed consent form, prior to any research procedures.

Social Cognitive Assessments. The Emotion Recognition (ER40) task consists of 40 images of whole faces making emotional expressions, with participants selecting from four possible emotional responses<sup>34</sup>. The Reading the Mind in the Eyes Test (RMET)<sup>35</sup> consists of 36 trials showing only the eyes of a black and white face, with four possible choices to describe what the person is feeling (e.g. amused, irritated, cautious, contemplative). The RMET is considered a test of empathic abilities as it measures the ability to judge emotional states from looking at the eyes. The Relationships Across Domains (RAD) test<sup>36</sup> presents 25 written vignettes of 2-4 lines followed by three statements which describe the behaviour of the male-female dyad from each vignette in domains of social life different from that vignette. Participants indicate if the behavior described in each statement is likely or unlikely to occur based on what was learned from the vignettes. The Awareness of Social Inference Test, Revised (TASIT) uses short video vignettes to measure emotional perception and theory of mind<sup>37</sup>. The TASIT is divided into three parts. Part 1 (TASIT1) consists of 24 short videos of actors portraying different emotional states (happy, sad, fear, disgust, surprise, and anger). Part 2 (TASIT2; social inference - minimal) consists of 15 videos showing sincere or sarcastic interactions between two actors, followed by four questions relating to what the actors were thinking, doing, meaning to say, and feeling. TASIT2 is considered a test of theory of mind. TASIT part 3 (TASIT3; social inference - enriched) consists of 15 vignettes in which the speaker is making an assertion which is literally untrue, but can represent either sarcasm or an attempt at deception. Success in TASIT3 requires the ability to detect deception in social encounters. In total, six social cognitive scores were derived from these tests (ER40 reaction time, RMET, RAD TASIT1, TASIT2, and TASIT3). Age was regressed out of all social cognitive test scores.

**MRI Scanning.** MRI scanning was conducted on a Discovery 3 T MR750 machine from General Electric at the Centre for Addiction and Mental Health. The Imitate/Observe task was part of a longer multimodal MRI protocol to which each participant consented. Each block of the task (Imitate and Observe) was collected in counter-balanced order as a separate echo-planar imaging scan, with TRs = 3 sec, TE 30 ms, voxel size 3 mm isotropic, 50 slices,  $64 \times 64$  matrix with FOV = 192 mm, flip angle = 77, and 110 TRs per scan. The first 5 TRs were excluded prior to any preprocessing to allow for magnetic steady-state.

**Imitate/Observe Task.** The participants performed an imitate/observe task as previously described<sup>13-15</sup> while being scanned. Participants were shown full-color photographs from an ethnically diverse set of 16 individuals (eight males and eight females) expressing five different facial expressions (fearful, sad, happy, angry, or neutral). During one scanning session, participants were instructed to imitate the expression on the faces (the *Imitate* session), while in the other session participants were instructed to observe the face (the *Observe* session). Each run consisted of 80 faces (16 per facial expression) plus 16 fixation trials presented in a pseudorandomized order determined using Optimize Design 11<sup>67</sup> to maximize contrast efficiency. Each trial lasted three seconds, with the faces presented for two seconds and a jittered ISI (500 ms to 1500 ms). The order of sessions (Imitate or Observe) was counter-balanced across participants. Participants practiced the task prior to MRI scanning and were instructed to minimize motion during the imitation, only using their facial muscles when matching the expressions. A video recording during the scan confirmed that all participants were following instructions (i.e. imitating during the imitate session but not during observe).

**fMRI data analysis - Preprocessing.** The initial preprocessing stages were slice time corrected and motion corrected in SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Individual sessions were subjected to single-session independent components analysis (ICA, from the MELODIC module in FSL 5.0.6). ICA reports were visually examined, and any ICAs which were clear artifacts were removed based on a set of established criteria. On a case by case basis where some large motions contaminated the ICAs, data 'scrubbing' was performed using a spline interpolation. Generally, two TRs were removed for each motion spike, and no more than three TRs were permitted per motion spike and a maximum of three motion spikes were removed from any scan. Of the 56 sessions (Observe and Imitate for the 28 participants) scrubbing was performed in 16 sessions. In nine of these cases (four Observe sessions and five Imitate sessions) scrubbing was not successful. Possible reasons for unsuccessful 'data scrubbing' include large movements exceeding three TRs, more than three large

movements, or over 90% of ICA components classified as artifact after data scrubbing. These sessions were subsequently excluded from further analysis, leaving neuroimaging data from 20 of the original 28 individuals for final analysis. Data were then de-noised based on the selected noise ICA components using FSL 5.0.6 (reg\_filt). Normalization into MNI space was done using SPM8, and images were smoothed with an 8 mm Gaussian kernel.

**Partial Least Squares Analysis.** Patterns of task related activity and relationships between social cognitive test scores and BOLD signal were examined using spatio-temporal PLS<sup>31, 32</sup>, a multivariate approach which allows for detection of spatial patterns and dependant variables across the brain without the need for a-priori selected contrasts across experimental conditions. PLS is designed to handle data with high collinearity while capturing essential non-redundant relationships among the data<sup>31</sup>, overcoming the confounding influence of collinear variables (e.g. cognitive tests) in traditional multiple regression. Additionally, PLS is also well suited to studying the relationships between numerous variables even in the presence of a relatively small sample, which is ideal for our purposes as it allows us to examine a range of social cognitive batteries. PLS produces latent variables (LVs) relating patterns of experiment task activity or behavioral measures with spatial patterns of neural activity across time points (scans, referred to as *lags*, normalized to the scan in which the stimuli was presented, labeled as lag0). As a model free non-parametric approach, PLS is ideal to examine complex relationships amongst the battery of social cognitive scores and neural activity when imitating and observing emotional faces.

A task-PLS analysis was conducted to replicate the regions of activity in the SPM8 GLM. Brain data from an 18 second window (corresponding to 6 lags) was normalized to the first lag (lag0) to create a data matrix for each condition, stacked across participants. Cross covariance was calculated between the data matrix and a design matrix consisting of vector of experimental conditions (in the task PLS, emotional faces, neutral faces, and fix-ation crosses, separately for Imitate and Observe). The resulting cross-covariance matrix was then decomposed using singular value decomposition, which created a set of orthogonal latent variables (**LVs**) which optimally represent spatio-temporal relationships between voxels and experimental conditions. For each LV, a pattern of voxels at each lag value (the 'brain pattern') demonstrates the relationship with activity in the voxel at that lag and the 'design pattern' (representing the weights of each experimental event). Voxel weight is expressed as salience, which is proportional to the covariance of activity in that voxel and the design pattern expressed by that LV.

Behavioral PLS was used to examine relationships between social cognition and neural activity. Behavioral PLS examines patterns of covariance between scores and neural activity for each trial type across lags. As such the design pattern is the correlation between the overall pattern of neural activity in each condition and each social cognitive score. For the behavioral PLS, neural activity to emotional faces, neutral faces, and fixation (separately for imitate and observe) was related to the six social cognitive test scores from our battery, deterministically producing 36 LVs.

Statistical evaluation of each LV was performed using split-half permutation testing. A total of 500 permutations were run, with 100 split-half permutations within each. The overall permutation score determines if the effect represented by the LV is sufficiently strong to be differentiated from random noise, while the split-half analysis provides a measure of the stability of relationships between voxel patterns and design patterns in the data for each latent variable. As the permutations are performed at the level of the entire PLS analysis (rather than on individual LVs), multiple comparisons for the number of LVs is not necessary. In the split-half, the 'BrainCorr' value tests the reliability of the voxel pattern for a given design pattern associated with an LV, while 'DesignCorr' tests the reliability of a given voxel pattern for the behavioral pattern associated with that LV. Results are expressed as p-values. LV's were considered significant if the overall permutations was used to test if specific voxels were reliably related to the LV. A bootstrap ratio for each voxel was calculated as the voxel salience divided by its bootstrap standard error. A bootstrap ratio of 2.5 (corresponding to >95% reliability) was used to threshold all voxel pattern maps in PLS.

#### References

- Mancuso, F., Horan, W. P., Kern, R. S. & Green, M. F. Social cognition in psychosis: multidimensional structure, clinical correlates, and relationship with functional outcome. *Schizophrenia research* 125, 143–151, doi:10.1016/j.schres.2010.11.007 (2011).
- Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. NeuroImage 48, 564–584, doi:10.1016/j.neuroimage.2009.06.009 (2009).
- Herbet, G. et al. Inferring a dual-stream model of mentalizing from associative white matter fibres disconnection. Brain: a journal of neurology 137, 944–959, doi:10.1093/brain/awt370 (2014).
- Shamay-Tsoory, S. G., Aharon-Peretz, J. & Perry, D. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: a journal of neurology* 132, 617–627, doi:10.1093/brain/awn279 (2009).
- 5. Caspers, S., Zilles, K., Laird, A. R. & Eickhoff, S. B. ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage* **50**, 1148–1167, doi:10.1016/j.neuroimage.2009.12.112 (2010).
- Coricelli, G. Two-levels of mental states attribution: from automaticity to voluntariness. *Neuropsychologia* 43, 294–300, doi:10.1016/j.neuropsychologia.2004.11.015 (2005).
- Spunt, R. P. & Lieberman, M. D. The busy social brain: evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychological science* 24, 80–86, doi:10.1177/0956797612450884 (2013).
- Spunt, R. P. & Lieberman, M. D. Dissociating modality-specific and supramodal neural systems for action understanding. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 32, 3575–3583, doi:10.1523/JNEUROSCI.5715-11.2012 (2012).
- Tunik, E., Rice, N. J., Hamilton, A. & Grafton, S. T. Beyond grasping: representation of action in human anterior intraparietal sulcus. NeuroImage 36(Suppl 2), T77–86, doi:10.1016/j.neuroimage.2007.03.026 (2007).
- Gallese, V. & Goldman, A. Mirror neurons and the simulation theory of mind-reading. Trends in cognitive sciences 2, 493-501, doi:10.1016/S1364-6613(98)01262-5 (1998).
- 11. Eisenberg, N. Emotion, regulation, and moral development. *Annual review of psychology* **51**, 665–697, doi:10.1146/annurev. psych.51.1.665 (2000).

- Tangney, J. P., Stuewig, J. & Mashek, D. J. Moral emotions and moral behavior. Annual review of psychology 58, 345–372, doi:10.1146/ annurev.psych.56.091103.070145 (2007).
- 13. Dapretto, M. *et al.* Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience* **9**, 28–30, doi:10.1038/nn1611 (2006).
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C. & Lenzi, G. L. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America* 100, 5497–5502, doi:10.1073/pnas.0935845100 (2003).
- Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C. & Dapretto, M. Mirroring others' emotions relates to empathy and interpersonal competence in children. *NeuroImage* 39, 2076–2085, doi:10.1016/j.neuroimage.2007.10.032 (2008).
- Vogeley, K. et al. Mind reading: neural mechanisms of theory of mind and self-perspective. NeuroImage 14, 170–181, doi:10.1006/ nimg.2001.0789 (2001).
- Samson, D., Apperly, I. A., Kathirgamanathan, U. & Humphreys, G. W. Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain: a journal of neurology* 128, 1102–1111, doi:10.1093/brain/awh464 (2005).
- Voineskos, A. N. et al. Neuroimaging evidence for the deficit subtype of schizophrenia. JAMA psychiatry 70, 472–480, doi:10.1001/jamapsychiatry.2013.786 (2013).
- 19. Wheeler, A. L. et al. Disrupted prefrontal interhemispheric structural coupling in schizophrenia related to working memory performance. Schizophrenia bulletin 40, 914–924, doi:10.1093/schbul/sbt100 (2014).
- Bliksted, V., Fagerlund, B., Weed, E., Frith, C. & Videbech, P. Social cognition and neurocognitive deficits in first-episode schizophrenia. Schizophrenia research 153, 9–17, doi:10.1016/j.schres.2014.01.010 (2014).
- Mikulan, E. P., Reynaldo, L. & Ibanez, A. Homuncular mirrors: misunderstanding causality in embodied cognition. Frontiers in human neuroscience 8, 299, doi:10.3389/fnhum.2014.00299 (2014).
- Caramazza, A., Anzellotti, S., Strnad, L. & Lingnau, A. Embodied cognition and mirror neurons: a critical assessment. Annual review of neuroscience 37, 1–15, doi:10.1146/annurev-neuro-071013-013950 (2014).
- Corradini, A. & Antonietti, A. Mirror neurons and their function in cognitively understood empathy. *Consciousness and cognition* 22, 1152–1161, doi:10.1016/j.concog.2013.03.003 (2013).
- Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G. & Walter, H. Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. Social cognitive and affective neuroscience 9, 909–916, doi:10.1093/scan/nst062 (2014).
- Wolf, I., Dziobek, I. & Heekeren, H. R. Neural correlates of social cognition in naturalistic settings: a model-free analysis approach. NeuroImage 49, 894–904, doi:10.1016/j.neuroimage.2009.08.060 (2010).
- Zaki, J., Hennigan, K., Weber, J. & Ochsner, K. N. Social cognitive conflict resolution: contributions of domain-general and domainspecific neural systems. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 30, 8481–8488, doi:10.1523/ JNEUROSCI.0382-10.2010 (2010).
- Zaki, J., Weber, J., Bolger, N. & Ochsner, K. The neural bases of empathic accuracy. Proceedings of the National Academy of Sciences of the United States of America 106, 11382–11387, doi:10.1073/pnas.0902666106 (2009).
- Marsh, L. E. & Hamilton, A. F. Dissociation of mirroring and mentalising systems in autism. *NeuroImage* 56, 1511–1519, doi:10.1016/j.neuroimage.2011.02.003 (2011).
- Marsh, L. E., Mullett, T. L., Ropar, D. & Hamilton, A. F. Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage* 103, 81–90, doi:10.1016/j.neuroimage.2014.09.020 (2014).
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I. & Bekkering, H. Complementary systems for understanding action intentions. *Current biology: CB* 18, 454–457, doi:10.1016/j.cub.2008.02.057 (2008).
- McIntosh, A. R. & Lobaugh, N. J. Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* 23(Suppl 1), S250–263, doi:10.1016/j.neuroimage.2004.07.020 (2004).
- 32. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage* **56**, 455–475, doi:10.1016/j.neuroimage.2010.07.034 (2011).
- Wold, S., Ruhe, A., Wold, H. & Dunn, I. W. J. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical Computing 5, 735–743, doi:10.1137/0905052 (1984).
- Kohler, C. G., Bilker, W., Hagendoorn, M., Gur, R. E. & Gur, R. C. Emotion recognition deficit in schizophrenia: association with symptomatology and cognition. *Biological psychiatry* 48, 127–136, doi:10.1016/S0006-3223(00)00847-7 (2000).
- 35. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry, and allied disciplines* 42, 241–251, doi:10.1111/jcpp.2001.42.issue-2 (2001).
- Sergi, M. J. et al. Development of a measure of relationship perception in schizophrenia. Psychiatry research 166, 54–62, doi:10.1016/j.psychres.2008.03.010 (2009).
- 37. McDonald, S., Flanagan, S. & Rollins, J. The Awareness of Social Inference Test Revised (TASIT-R) (Pearson Assessment, 2011).
- Trapp, K. et al. Imagining triadic interactions simultaneously activates mirror and mentalizing systems. NeuroImage 98, 314–323, doi:10.1016/j.neuroimage.2014.05.003 (2014).
- Spengler, S., Bird, G. & Brass, M. Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological psychiatry* 68, 1148–1155, doi:10.1016/j.biopsych.2010.09.017 (2010).
- Brass, M., Ruby, P. & Spengler, S. Inhibition of imitative behaviour and social cognition. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 2359–2367, doi:10.1098/rstb.2009.0066 (2009).
- 41. Rizzolatti, G. & Sinigaglia, C. Further reflections on how we interpret the actions of others. *Nature* 455, 589-589, doi:10.1038/455589b (2008).
- Casartelli, L. & Molteni, M. Where there is a goal, there is a way: what, why and how the parieto-frontal mirror network can mediate imitative behaviours. *Neuroscience and biobehavioral reviews* 47, 177–193, doi:10.1016/j.neubiorev.2014.08.004 (2014).
- de Bruin, L. & Gallagher, S. Embodied simulation, an unproductive explanation: comment on Gallese and Sinigaglia. Trends in cognitive sciences 16, 98–99, doi:10.1016/j.tics.2011.12.003 (2012). author reply 99-100.
- 44. Gallese, V. & Sinigaglia, C. What is so special about embodied simulation? *Trends in cognitive sciences* 15, 512–519, doi:10.1016/j. tics.2011.09.003 (2011).
- Agnew, Z. K., Bhakoo, K. K. & Puri, B. K. The human mirror system: a motor resonance theory of mind-reading. Brain research reviews 54, 286–293, doi:10.1016/j.brainresrev.2007.04.003 (2007).
- 46. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Human brain mapping* **30**, 829–858, doi:10.1002/hbm.20547 (2009).
- Zaki, J. & Ochsner, K. N. The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience* 15, 675–680, doi:10.1038/ nn.3085 (2012).
- Ochsner, K. N. The social-emotional processing stream: five core constructs and their translational potential for schizophrenia and beyond. *Biological psychiatry* 64, 48–61, doi:10.1016/j.biopsych.2008.04.024 (2008).
- Lisofsky, N., Kazzer, P., Heekeren, H. R. & Prehn, K. Investigating socio-cognitive processes in deception: a quantitative metaanalysis of neuroimaging studies. *Neuropsychologia* 61, 113–122, doi:10.1016/j.neuropsychologia.2014.06.001 (2014).
- Iacoboni, M. & Dapretto, M. The mirror neuron system and the consequences of its dysfunction. *Nature reviews. Neuroscience* 7, 942–951, doi:10.1038/nrn2024 (2006).

- Yang, D. Y., Rosenblau, G., Keifer, C. & Pelphrey, K. A. An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience and biobehavioral reviews* 51, 263–275, doi:10.1016/j.neubiorev.2015.01.020 (2015).
- Nelissen, K. et al. Action observation circuits in the macaque monkey cortex. The Journal of neuroscience: the official journal of the Society for Neuroscience 31, 3743–3756, doi:10.1523/JNEUROSCI.4803-10.2011 (2011).
- 53. Ishai, A. Let's face it: it's a cortical network. NeuroImage 40, 415-419, doi:10.1016/j.neuroimage.2007.10.040 (2008).
- 54. Ishai, A., Schmidt, C. F. & Boesiger, P. Face perception is mediated by a distributed cortical network. *Brain research bulletin* 67, 87–93, doi:10.1016/j.brainresbull.2005.05.027 (2005).
- Bonda, E., Petrides, M., Ostry, D. & Evans, A. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *The Journal of neuroscience: the official journal of the Society for Neuroscience* 16, 3737–3744 (1996).
- Grossman, E. et al. Brain areas involved in perception of biological motion. Journal of cognitive neuroscience 12, 711-720, doi:10.1162/089892900562417 (2000).
- Redcay, E. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience and biobehavioral reviews* 32, 123–142, doi:10.1016/j.neubiorev.2007.06.004 (2008).
- Zilbovicius, M. et al. Autism, the superior temporal sulcus and social perception. Trends in neurosciences 29, 359–366, doi:10.1016/j. tins.2006.06.004 (2006).
- Kana, R. K., Libero, L. E., Hu, C. P., Deshpande, H. D. & Colburn, J. S. Functional brain networks and white matter underlying theory-of-mind in autism. Social cognitive and affective neuroscience 9, 98–105, doi:10.1093/scan/nss106 (2014).
- Lombardo, M. V. et al. Shared neural circuits for mentalizing about the self and others. Journal of cognitive neuroscience 22, 1623–1635, doi:10.1162/jocn.2009.21287 (2010).
- Sperduti, M., Guionnet, S., Fossati, P. & Nadel, J. Mirror Neuron System and Mentalizing System connect during online social interaction. *Cognitive processing* 15, 307–316, doi:10.1007/s10339-014-0600-x (2014).
- Spunt, R. P. & Lieberman, M. D. An integrative model of the neural systems supporting the comprehension of observed emotional behavior. *NeuroImage* 59, 3050–3059, doi:10.1016/j.neuroimage.2011.10.005 (2012).
- 63. Libero, L. E. *et al.* The role of mirroring and mentalizing networks in mediating action intentions in autism. *Molecular autism* 5, 50, doi:10.1186/2040-2392-5-50 (2014).
- Horan, W. P. et al. Self-reported empathy and neural activity during action imitation and observation in schizophrenia. NeuroImage. Clinical 5, 100–108, doi:10.1016/j.nicl.2014.06.006 (2014).
- Fishman, I., Keown, C. L., Lincoln, A. J., Pineda, J. A. & Muller, R. A. Atypical cross talk between mentalizing and mirror neuron networks in autism spectrum disorder. *JAMA psychiatry* 71, 751–760, doi:10.1001/jamapsychiatry.2014.83 (2014).
- 66. van Asselen, M. *et al.* Scanning Patterns of Faces do not Explain Impaired Emotion Recognition in Huntington Disease: Evidence for a High Level Mechanism. *Frontiers in psychology* 3, 31, doi:10.3389/fpsyg.2012.00031 (2012).
- 67. Wager, T. D. & Nichols, T. E. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* 18, 293–309, doi:10.1016/S1053-8119(02)00046-0 (2003).

#### Acknowledgements

C.H. was supported by a Banting Post-Doctoral Fellowship from the Canadian Institute for Health Research. M.I. supported by NIH grant MH097178. ANV is supported by the CAMH Foundation, Ontario Ministry of Research and Innovation, CIHR, Brain and Behavior Research Foundation, Canada Foundation for Innovation. National Institute of Mental Health R01MH102324, R01MH102313, and R01MH102318 supported the present work.

### **Author Contributions**

C.H. performed data analysis and prepared the manuscript. A.N.V. helped prepare the manuscript, managed data collection. A.V.N., A.K.M., and R.W.B. conceived of and designed the study. A.R.M. provided software, statistical consultation, and advice on data interpretation. N.K. assisted in data analysis. J.D.V. assisted in analysis and interpretation. M.I. provided the task, and contributed to interpretation of the results. All authors reviewed and contributed to the final manuscript.

### **Additional Information**

Competing Interests: The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017