

The neural dynamics of updating person impressions

Peter Mende-Siedlecki,¹ Yang Cai,¹ and Alexander Todorov^{1,2}

¹Department of Psychology, Princeton University, Princeton, NJ 08540, USA and ²Behavioural Science Institute, Radboud University, 6500 HE Nijmegen, the Netherlands

Person perception is a dynamic, evolving process. Because other people are an endless source of social information, people need to update their impressions of others based upon new information. We devised an fMRI study to identify brain regions involved in updating impressions. Participants saw faces paired with valenced behavioral information and were asked to form impressions of these individuals. Each face was seen five times in a row, each time with a different behavioral description. Critically, for half of the faces the behaviors were evaluatively consistent, while for the other half they were inconsistent. In line with prior work, dorsomedial prefrontal cortex (dmPFC) was associated with forming impressions of individuals based on behavioral information. More importantly, a whole-brain analysis revealed a network of other regions associated with updating impressions of individuals who exhibited evaluatively inconsistent behaviors, including rostralateral PFC, superior temporal sulcus, right inferior parietal lobule and posterior cingulate cortex.

Keywords: impression formation; social cognition; person perception; dmPFC; IPFC; fMRI

INTRODUCTION

Human social interaction is as informationally rich as it is ubiquitous. As we spend countless hours engaging with other humans, we form impressions of the people around us—in large part, in an attempt to predict behavior. However, our fellow interaction partners are not always so consistent. As such, social interaction requires continuous, flexible updating of our initial impressions in light of new information.

An abundance of research has accumulated on the neural bases of ‘first impressions’. Much of this research has focused on initial appraisals of other people based on facial characteristics like attractiveness and perceived trustworthiness (for meta-analysis, see Mende-Siedlecki *et al.*, 2011). First impressions based upon behavioral information have been extensively examined as well. This research has shown that our impressions of the people around us are powerfully influenced by the behaviors we come to associate with them (Todorov and Uleman, 2002; Bliss-Moreau *et al.*, 2008; Todorov and Olson, 2008). Behavior-based impression formation can lead to automatic inferences regarding character traits (Todorov and Uleman, 2003), and further, can be generalized to similar-looking others (Verosky and Todorov, 2010). Typically, in such studies, people represented by faces paired with negative behavioral information are subsequently evaluated as being less trustworthy, and people paired with positive information are subsequently evaluated as being more trustworthy (Todorov and Olson, 2008).

Recent neuroimaging work has sought to identify brain regions crucial for forming impressions of others based upon behavioral information. The primary region associated with such tasks is the dorsomedial prefrontal cortex (dmPFC; Mitchell *et al.*, 2004, 2005, 2006; Schiller *et al.*, 2009; Baron *et al.*, 2011; Cloutier *et al.*, 2011a). Dovetailing with these results, the dmPFC has also been previously implicated in the spontaneous retrieval of affective person knowledge about faces previously learned in the context of behavioral information (Gobbini *et al.*, 2004; Gobbini and Haxby, 2007; Todorov *et al.*, 2007).

Regions such as the posterior cingulate cortex (PCC; Schiller *et al.*, 2009; Freeman *et al.*, 2010; Cloutier *et al.*, 2011a;), amygdala (Schiller *et al.*, 2009; Baron *et al.*, 2011), superior temporal sulcus (STS; Mitchell *et al.*, 2005; Schiller *et al.*, 2009; Freeman *et al.*, 2010) and inferior frontal gyrus (IFG; Mitchell *et al.*, 2005; Schiller *et al.*, 2009; Baron *et al.*, 2011; Freeman *et al.*, 2010) have also been observed in conjunction with this type of impression formation task. However, while it is possible to speculate on a putative network of regions involved in impression formation, the preponderance of studies implicating the dmPFC in such tasks is undeniable.

Although there is a substantial body of research on first impressions, much less is known about how these impressions are updated. Impression formation is an ongoing process, and initial impressions must be updated on the basis of new, incoming information—which may be evaluatively inconsistent with previous impressions. Here, we explore a phenomenon we describe as impression updating—situations where new information learned about a target is evaluatively inconsistent with a previous impression, thus necessitating an update of that impression to account for the inconsistency.

Social psychology affords us a host of predictions regarding how person perception can be affected by such a turn of events (Reeder and Brewer, 1979; Fiske, 1980; Reeder and Spores, 1983; Skowronski and Carlston, 1987, 1989). Our impressions of other people may function as schemas that drive our expectancies of their future behavior (Fiske and Linville, 1980). When we are faced with information that is inconsistent with a given schema, we are forced to reassess our impression to account for the new information (Srull and Wyer, 1989). However, despite previous behavioral work, neuroimaging investigations of impression updating have just begun. Some recent research has addressed the neural dynamics of how initial impressions are updated by behavioral information, in both electrophysiological (Rudoy and Paller, 2009) and neuroimaging contexts (Baron *et al.*, 2011; Cloutier *et al.*, 2011b; Ma *et al.*, 2011). Baron and colleagues presented participants with untrustworthy-, trustworthy- and neutral-looking faces in the scanner, and in a subsequent phase, paired some of these faces with valenced behavioral information. Not only was the dmPFC more active during learning for faces paired with behaviors, but this activity correlated with a post-scan measure of learning, suggesting that in the context of this task, the dmPFC plays an important role in updating initial appearance-based impressions based upon behavioral information.

Received 24 January 2012; Accepted 17 March 2012
Advance Access publication 3 April 2012

This research was supported by National Science Foundation grant BCS-0823749 (to A.T.), National Science Foundation (grant DGE 1148900 to P.M.-S.), the Princeton University Class of '55 Senior Thesis Fund (to Y.C.) and the Russell Sage Foundation. The authors would also like to thank Jenny Porter and Hillel Aviezer for assistance in running participants and Tor Wager for making the MKDA Toolbox available online (which was used to make Figure 4). Finally, the authors thank Sara Verosky, Sean Baron, and Dan Ames for their helpful comments and advice.

Correspondence should be addressed to Peter Mende-Siedlecki, 2-N-9, Green Hall, Princeton University, Princeton, NJ, 08540. E-mail: pmende@princeton.edu

Especially relevant is a recent study by Ma and colleagues, in which participants read sets of behavioral descriptions that implied a specific trait about a particular individual. Critically, the last behavior was manipulated to be either consistent or inconsistent with that implied trait. Responses in the dmPFC were higher when this last behavior was trait-inconsistent, compared to when it was trait-consistent (Ma *et al.*, 2011). Finally, another recent study by Cloutier and colleagues observed preferentially higher dmPFC activity when targets' behaviors were incongruent with their social category (in this case, political affiliation), as opposed to when they were congruent (Cloutier *et al.*, 2011b).

The current study focuses on evaluative impression updating over a long behavioral trajectory. To that aim, we presented participants with person targets who were paired with five descriptions of valenced behaviors (e.g. 'Ron gave out toys at the children's hospital during Christmas'), viewed consecutively. Half of the targets were paired with behavioral information that remained either consistently negative or consistently positive, thus requiring little demand for impression updating. The other half of the targets were paired with behavioral information that switched valence on the fourth trial. The desired effect is that the first three pieces of behavioral information create a strong expectation for that person to behave in a certain manner (for instance, acting like a good, law-abiding citizen)—an expectation that is subsequently violated on trials four and five, resulting in a high demand for impression updating.

We expected that participants would update their impressions of targets based upon new, inconsistent information. More importantly, consistent with other studies (Mitchell *et al.*, 2004, 2005, 2006; Schiller *et al.*, 2009), we expected that evaluative updating of impressions would recruit regions implicated in impression formation such as the dmPFC. Finally, based on recent studies (Cloutier *et al.*, 2011b; Ma *et al.*, 2011), we expected that in addition to these regions, evaluative updating would recruit regions involved in attention and cognitive control.

METHODS

Participants

Twenty-four (14 female) participants volunteered for the fMRI study and were paid \$30 for their participation. They were between the ages of 18 and 45 years (mean = 25.3 years). All participants were right-handed, had normal or corrected-to-normal vision and reported no history of neurological illnesses or abnormalities. We acquired informed consent for participation approved by the Institutional Review Board for Human Subjects at Princeton University. All participants were fully debriefed at the completion of the experiment.

Face and behavior stimuli

Each participant saw a series of 50 faces taken from the book 'Heads' (Kayser, 1997), paired with positively and negatively valenced behaviors previously rated on goodness and kindness (Fuhrman *et al.*, 1989). Each face was paired with five consecutively viewed behaviors, comprising one 'target'. Targets were classified as either evaluatively consistent or inconsistent. Consistent targets consisted of a face paired with five behaviors of the same valence—either five straight positive behaviors (consistently positive) or five straight negative behaviors (consistently negative). Inconsistent targets consisted of a face paired with three behaviors of one valence, followed by two behaviors of the opposite valence—either three positive behaviors followed by two negative behaviors (positive-to-negative), or three negative behaviors followed by two positive behaviors (negative-to-positive). Additionally, participants sometimes saw control targets—faces presented alone on screen, without accompanying behaviors. All in all, participants

encountered 50 total targets—10 targets corresponding to each of these five conditions.

Behaviors were combined together in groups of five such that each group within a given condition would be roughly equated on goodness and kindness. The average goodness and kindness ratings for each condition were as follows: consistently negative ($M=1.81$, $SD=0.61$), negative-to-positive ($M=4.79$, $SD=3.15$), consistently positive ($M=8.10$, $SD=0.63$), positive-to-negative ($M=4.83$, $SD=3.20$). Faces and behavior valences were counterbalanced between participants, such that each face was paired with each type of behavior group an equal number of times. Finally, each participant was given a unique, optimized target ordering, based upon a genetic algorithm (Wager and Nichols, 2003, http://wagerlab.colorado.edu/wiki/doku.php/help/ga/genetic_algorithm_for_fmri) to maximize statistical power. We note that while facial trustworthiness is not of interest in this study, the faces we used indeed varied on this dimension. That said, due to the counterbalancing of faces and behavior valences, any differences due to facial trustworthiness are assumed to be negligible.

Procedures

Participants were informed that they would be participating in a study on impression formation. They were told that they would be seeing a series of faces paired with behaviors, and that they would see multiple behaviors paired consecutively with each face. Participants were asked to form an impression of each target, altering that impression if necessary based on new information they learned as the task went along. Additionally, participants were told that picturing targets performing behaviors would likely aid in forming impressions. In scanner, they saw ten runs of face targets, each paired with five separate behaviors.

Each run consisted of five face targets, one of each condition. Each run began with a 15 s presentation of a fixation cross. Each target was split into five face/behavior presentations. Faces and behaviors were presented together for 6 s. Next, a rating slide appeared for 4 s, during which the participant rated how trustworthy that individual was, based upon the information they had learned about him so far. Participants made their ratings with an MR-safe button box, on a scale ranging from 1 (very untrustworthy) to 4 (very trustworthy). Subsequently, a fixation cross appeared for 4 s. This series of events proceeded four more times per target (with the same face on the screen, paired with different behaviors each time). Following the fifth behavior, a new target appeared. All stimuli were projected onto a screen located at the rear of the bore of the magnet. Participants were able to view these stimuli via an angled mirror attached to the RF coil placed above their eyes.

Image acquisition

Blood oxygenation level-dependent (BOLD) signal was used as a measure of neural activation. Echo planar images (EPI) were acquired using a Siemens 3.0 Tesla Allegra head-dedicated scanner (Siemens, Erlangen, Germany) with a standard 'bird-cage' head coil ($TR=2000$ ms, $TE=30$ ms, flip angle = 80° , matrix size = 64×64). By using 32 interleaved 3-mm axial slices we were able to achieve near whole brain coverage. Prior to the primary data acquisition scan, a high-resolution anatomical image (T1-MPRAGE, $TR=2500$ ms, $TE=4.3$ ms, flip angle = 8° , matrix size = 256×256) was acquired for subsequent registration of functional activity to the participant's anatomy and for spatially normalizing data across participants.

Image analysis

All fMRI data were analyzed with Analysis of Functional NeuroImages software (AFNI; Cox, 1996). The first four EPI images from each run

were discarded to allow the MR signal to reach steady-state equilibrium. Participants' motion was corrected using a six-parameter 3D motion-correction algorithm following slice scan-time correction. Transient spikes were removed from the signal using the AFNI program 3dDespike. Subsequently, data were low-passed filtered with a frequency cut-off of 0.1 Hz following spatial smoothing with a 6-mm full width at half maximum (FWHM) Gaussian kernel. The signal was then normalized to percent signal change from the mean.

To identify regions that were more active when participants were forming impressions based on behaviors, we contrasted trials in which faces were paired with behaviors and trials in which faces were presented alone. This contrast yielded functional regions of interest (fROIs) involved in learning to associate behavioral information with faces, and by extension, forming behavior-based impressions of those person targets. We subsequently analyzed the parameter estimates in these fROIs as a function of the order of the behaviors (the first three vs the last two behaviors) and the evaluative consistency of the behaviors. Given the large number of fROIs yielded by the contrast of faces paired with behaviors and faces alone, the parametric map was thresholded at $\alpha = 0.0001$ (uncorrected). Furthermore, to select a minimum cluster size for corrected significance ($P < 0.05$), we performed a Monte Carlo simulation of null-hypothesis data, using the AlphaSim program included in the AFNI package. The Monte Carlo simulation indicated that a minimum cluster size of 8 voxels was appropriate.

To generate parameter estimates, we performed voxel-wise multiple regression on each participant's preprocessed imaging data. Twenty-five regressors of interest (five 6000-ms trials per target \times 5 types of target) were convolved with a canonical hemodynamic response function and entered into our general linear model (GLM). Additionally, we included several regressors of no interest, including head motion estimates and time points representing rating slide presentations. Each participant's parameter estimate maps were projected into Talairach space (Talairach and Tournoux, 1988) prior to performing any group-level analyses.

In addition to the fROI analyses, we performed a whole-brain analysis testing the interaction between trial number (last two trials vs first three trials) and evaluative consistency (consistent vs inconsistent). Finally, we performed separate whole-brain analyses contrasting the last two trials against the first three trials, in both consistent and inconsistent targets. Because we did not find reliable main effects of the valence of the behaviors and higher order interactions with this valence, we do not report analyses related to valence. However, we provide supplemental figures including the valence of the behaviors. All whole-brain analyses are reported using the same thresholding procedures as described above ($P < 0.05$ FDR-corrected; voxel-wise threshold, $P < 0.005$; minimum cluster-size threshold, 31 voxels).

RESULTS

Behavioral results

Because we were primarily interested in updating impressions, we focus on the changes in ratings in response to evaluatively inconsistent information. We computed separate averages across the first three and last two behaviors, isolating participants' evaluations of our targets before and after the potential introduction of evaluatively inconsistent information. We further subtracted the ratings of control targets (faces presented without behavioral information) from the consistent and inconsistent targets' ratings and recorded the absolute deviation from the control condition. These deviations provide a measure of the change in target evaluation. [See Supplementary Figure 1 for the means across all 5 (target type) \times 5 (trial number) conditions].

Participants updated their impressions of person targets based upon evaluatively inconsistent information. Specifically, the change in

participants' ratings from the first three to the last two behaviors was greater for inconsistent targets than consistent targets. A 2 (trial number: first three behaviors vs last two behaviors) \times 2 (consistency: consistent targets vs inconsistent targets) ANOVA revealed significant main effects of trial number [$F(1,23) = 13.37$, $P < 0.001$] and consistency [$F(1,23) = 89.52$, $P < 0.001$]. Critically, we observed a significant interaction between trial number and consistency [$F(1,23) = 69.92$, $P < 0.001$], such that the absolute deviation in trustworthiness ratings from the first three to the last two behaviors was greater for inconsistent targets ($M = 0.58$, $SE = 0.08$) than for consistent targets ($M = 0.29$, $SE = 0.04$).

The mean response time across trials was 1119.41 ms ($SE = 47.75$). To test for potential differences in difficulty in processing information about consistent and inconsistent targets, we submitted the response times to a 2 (trial number: first three behaviors vs last two behaviors) \times 2 (consistency: consistent targets vs inconsistent targets) ANOVA. Neither main effect was significant, nor was the interaction between trial number and consistency. Nevertheless, we also tested for simple effects, and observed that the effect of trial number was not significant for either consistent [$t(23) = 0.18$, $P = 0.858$] or inconsistent targets [$t(23) = -1.48$, $p = 0.153$].

fMRI results

Brain activity associated with impression formation

We contrasted face-plus-behavior trials against face-alone trials. This method of localizing fROIs associated with forming impressions of person targets based on behavioral information is consistent with previous research (Schiller *et al.*, 2009; Baron *et al.*, 2011).

We observed 13 fROIs that responded more strongly to faces paired with behavioral information than to faces presented alone (Supplementary Table 1). We next tested which fROIs responded to the introduction of new behavioral information *inconsistent* with prior impressions, looking for a specific pattern of response, such that activity remained consistent or dropped from the first three trials (F3) to the last two trials (L2) for consistent and control targets, but increased for inconsistent targets.

The only fROI that produced this pattern of response was the dmPFC. As shown in Figure 1, activity increased in response to inconsistent information, but decreased when information was consistent. We performed a 3 (target type: inconsistent, consistent, control) \times 2 (trial number: first three trials vs last two trials) repeated-measures ANOVA on the β values extracted from this fROI, observing an interaction between consistency and trial number [$F(2,46) = 5.45$, $P = 0.008$, $\eta^2 = 0.19$]. Splitting these analyses by target type, we observed that dmPFC signal rose from the first three trials to the last two trials for inconsistent targets [$F(1,23) = 24.67$, $P < 0.001$, $\eta^2 = 0.52$]. Conversely, dmPFC signal change was not significant for consistent [$F(1,23) = 1.21$, $P = 0.283$, $\eta^2 = 0.05$] or control targets [$F(1,23) = 0.934$, $P = 0.344$, $\eta^2 = 0.04$] (See Supplementary Figure 2 for expanded analyses split by valence).

Brain activity associated with updating impressions

Interaction analysis. We sought to identify brain areas that showed a stronger L2 > F3 pattern for inconsistent targets than consistent targets, potentially reflecting their role in updating impressions based upon new, conflicting information. This interaction analysis showed that right IPL, left STS, PCC extending into the pulvinar, and bilateral rLPFC were all significantly more active in the last two trials than the first three trials for inconsistent targets only (Table 1 and Figure 2). In addition, right STS showed a similar pattern, though this cluster did not surpass extent-based thresholding. Visualizations of signal change

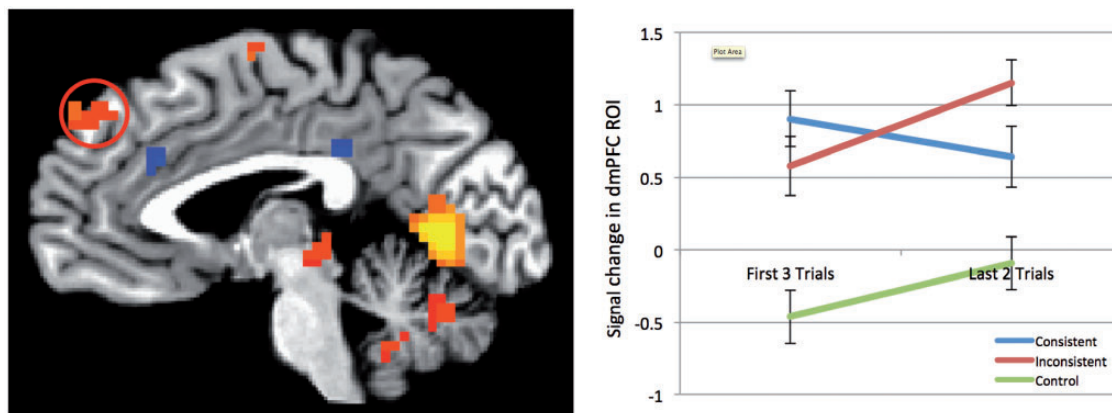


Fig. 1 Parameter estimates from dmPFC ROI from the Faces + Behaviors > Faces Alone contrast, split by evaluative consistency. Hot activations represent stronger activation for Faces + Behaviors, cold activations represent stronger activation for Faces Alone. While activity in the dmPFC (indicated by circle) did not change significantly from the first three to the last two trials in consistent targets, there was a significant increase in dmPFC activity from the first three to the last two trials in inconsistent targets.

in these regions are provided in Figure 2 (See Supplementary Figure 3 for expanded analyses split by valence).

L2 > F3 analyses, split by target type. To supplement the results of the interaction analysis, we performed separate L2 > F3 analyses for both consistent and inconsistent targets. Within consistent targets, we observed no brain areas that were preferentially active during the last two trials, while bilateral fusiform gyrus, cuneus and right pulvinar were more active during the first three trials (Supplementary Table 2, Figure 3).

However, the L2 > F3 contrast within inconsistent targets yielded activity in dmPFC, PCC/precuneus, bilateral rIPFC, bilateral dIPFC, bilateral IPL, bilateral STS and left anterior insula (Supplementary Table 2, Figure 3). The reverse contrast, F3 > L2, yielded activity in bilateral fusiform, cerebellum, right lingual gyrus, and inferior occipital gyrus.

DISCUSSION

To explore the neural dynamics of updating person impressions, we presented participants with faces paired with behavioral descriptions that were either consistent or inconsistent in valence. As expected, forming impressions of these targets based upon behavioral information, compared to presentation of faces alone, activated a set of regions typically associated with similar impression formation tasks, including the dmPFC. Within this set of regions, only the dmPFC showed preferential activation to updating based on new, evaluatively inconsistent information, as opposed to updating based on information consistent with existing impressions. Additional whole-brain analyses pointed to a larger set of regions involved in updating of evaluative impressions, including bilateral rIPFC, bilateral STS, PCC and right IPL.

We also observed regions that did not respond differentially as a function of the evaluative consistency of the behaviors. Specifically, large portions of inferotemporal cortex, including the bilateral fusiform gyri, were less active for the last two trials than the first three trials for both consistent and inconsistent targets (Figure 3), most likely a result of habituation in response to the repeatedly-presented facial stimuli (Kanwisher and Yovel, 2006).

The role of dmPFC in impression updating

The results of the fROI analyses showed that the dmPFC was the only region that displayed enhanced responses to evaluatively inconsistent but not to evaluatively consistent information, suggesting that it plays

an integral role in the evaluative updating of person impressions. This is consistent with previous conceptualizations of the dmPFC's role in impression formation (Mitchell *et al.*, 2004; 2005; 2006; Schiller *et al.*, 2009; Baron *et al.*, 2011; Ma *et al.*, 2011). Surprisingly, the whole brain interaction analysis of evaluative consistency and order of behaviors only yielded sub-threshold dmPFC activity—a discrepancy most likely due to the low-power nature of our design. In fact, the simple contrast comparing the last two *vs* first three behaviors *did* yield a large dmPFC activation for inconsistent but not consistent targets (Figure 3).

Two recent studies have also linked the dmPFC to impression updating. Ma and colleagues observed increased dmPFC activity in response to targets that behaved in a manner inconsistent with specific traits they had been previously associated with (Ma *et al.*, 2011). In addition, Cloutier and colleagues observed that the dmPFC also responded preferentially to instances where targets' behaviors were inconsistent with their social category (e.g. a Democrat favoring small government). In the context of this recent research, the present study suggests that the dmPFC's role in updating extends more broadly into instances of general evaluative inconsistency as well.

An alternative explanation of the increased dmPFC activity for inconsistent targets is that presenting inconsistent information on screen resulted in a less fluent reading experience. Hence, the increase in dmPFC activity is indicative of an increased difficulty associated with these targets. However, we observed no significant differences in response times across the last two trials between consistent and inconsistent targets, suggesting that our imaging results cannot be simply explained in terms of task difficulty.

A functional network for updating impressions

We now turn our attention to the other regions implicated in by our analyses. How might the STS, IPL, rIPFC and PCC be acting in service of impression updating? The STS has been previously demonstrated to play an integral role in a variety of tasks associated broadly with social processing and social cognition (Hein and Knight, 2008). Neuroimaging research in the past decade has frequently implicated the STS in aspects of high-level person perception critical for social communication, for instance, biological motion (Allison *et al.*, 2000; Vaina *et al.*, 2001; Grossman and Blake, 2002; Pelphrey *et al.*, 2003a; Puce and Perrett, 2003; Pelphrey *et al.*, 2004a; Pelphrey *et al.*, 2006) and facial expressions (static: Haxby *et al.*, 2000; Hoffman and Haxby, 2000; Adolphs, 2002; LaBar *et al.*, 2003; Calder and Young, 2005;

Table 1 Regions showing significant differences in the interaction contrast of last two trials vs first three trials as a function of consistency

Region	Lat	x	y	z	#Voxels
Interaction between $L2 > F3_{\text{Inconsistent}}$ and $L2 > F3_{\text{Consistent}}$					
Inferior parietal lobule	R	46.5	−64.5	47.5	317
PCC/pulvinar	—	1.5	−31.5	8.5	116
STS	L	−67.5	28.5	−2.5	86
Rostrolateral PFC	R	43.5	55.5	2.5	60
Rostrolateral PFC	L	−46.5	52.5	2.5	40
STS	R	64.5	−34.5	−9.5	28 ^a

All clusters are significant at $P < 0.05$, after correction for multiple comparisons, unless indicated with an asterisk. x, y, z coordinates reflect peak voxel location in Talairach coordinate system.
^aDid not surpass cluster extent-thresholding ($k = 31$).

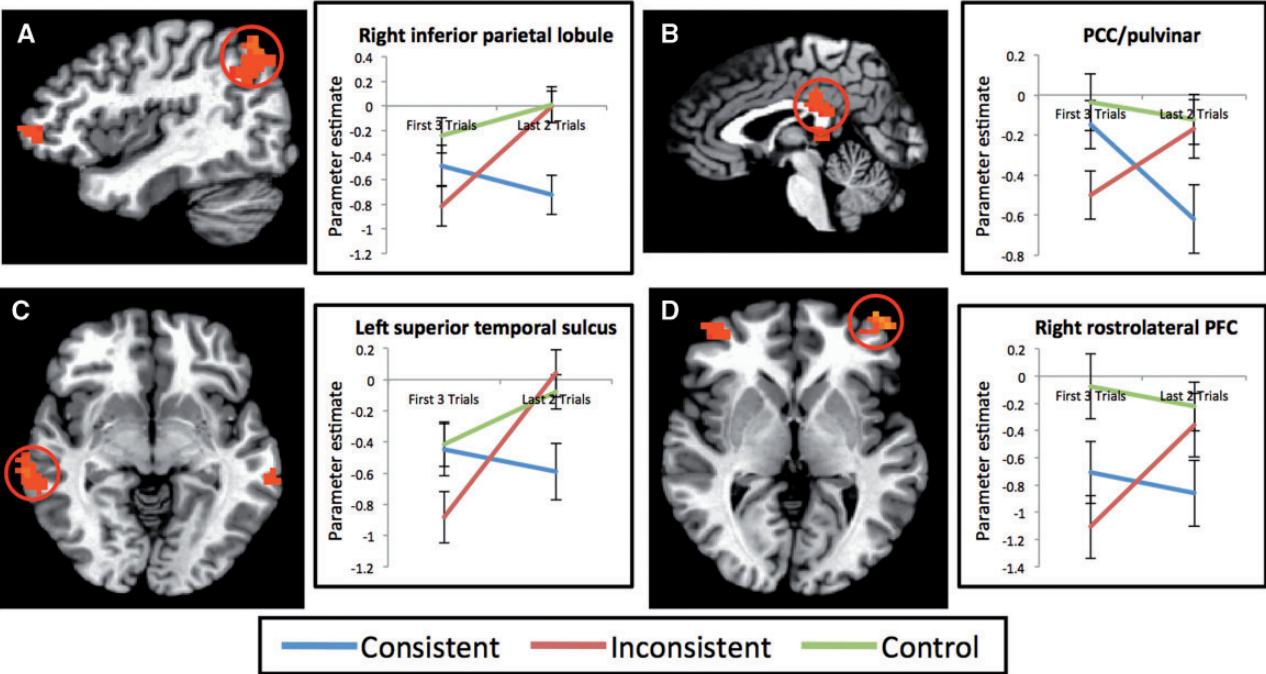


Fig. 2 Parameter estimates from regions of interest emerging from the interaction analysis between trial number and evaluative consistency. Hot activations indicate preferentially higher responses to the last two trials compared to the first three trials of each behavioral sequence, but only for inconsistent targets. Right IPL (A), PCC (B), left STS (C) and right rPFC (D) all showed a similar pattern, in which activity increased across the last two trials for inconsistent targets, but decreased for control targets.

Engell and Haxby, 2007; Ishai, 2008; dynamic: Ghazanfar *et al.*, 2010; Said *et al.*, 2010).

Meanwhile, the IPL has also been associated with a range of social cognitive functions, including gaze processing (Wicker *et al.*, 1998; Pelphrey *et al.*, 2003b; Pelphrey *et al.*, 2004b; Calder *et al.*, 2007), imitation (Iacoboni *et al.*, 1999; Decety *et al.*, 2002; Leslie *et al.*, 2004), action perception in the service of understanding intentions (Gallese *et al.*, 2004; Fogassi *et al.*, 2005; Iacoboni *et al.*, 2005; Montgomery and Haxby, 2008), self-other distinctions (Ruby and Decety, 2001; Ruby and Decety, 2003; Uddin *et al.*, 2006) and shared representations (Keysers *et al.*, 2004; Zaki *et al.*, 2009).

Many of the functions listed above are inherently germane to impression updating. First and foremost, both the STS and IPL have been connected to aspects of face processing. The omnipresence of facial stimuli in our task certainly introduces a prevalent, if implicit demand to process facial features. Furthermore, as we told our participants that they should imagine targets performing the actions they were paired with, it is possibly not surprising that an area like the IPL, associated with action perception (especially social actions), should be implicated.

Of most relevance, a recent review of research on the social brain suggests that one function of the STS is to predict the behavior of social agents based on incoming information (Frith and Frith, 2010). Specifically, the authors offer evidence suggesting that activity in posterior STS increases when a social agent behaves in a manner that is inconsistent with prior expectancies. In previous research, this inconsistency has taken the form of unexpected shifts in gaze (Pelphrey *et al.*, 2003b; Pelphrey *et al.*, 2004a), as well as unexpected changes in actions (Saxe *et al.*, 2004). In this sense, posterior STS activity in these tasks may be representing a social prediction error signal. Behrens and colleagues (2008) sought to directly test this possibility in a task in which participants made decisions based, in part, on a confederate's advice. This advice was occasionally unexpectedly incorrect or correct, eliciting a prediction error correlating with an increase in posterior STS activity, a signal dissociable from reward-related non-social prediction error signals observed in the ventral striatum. The results of the present study are consistent with this framework. On trials when evaluatively inconsistent information was presented, our participants' expectations were violated, and in turn, they were faced with the task of updating their impressions in order to better predict targets' future actions.

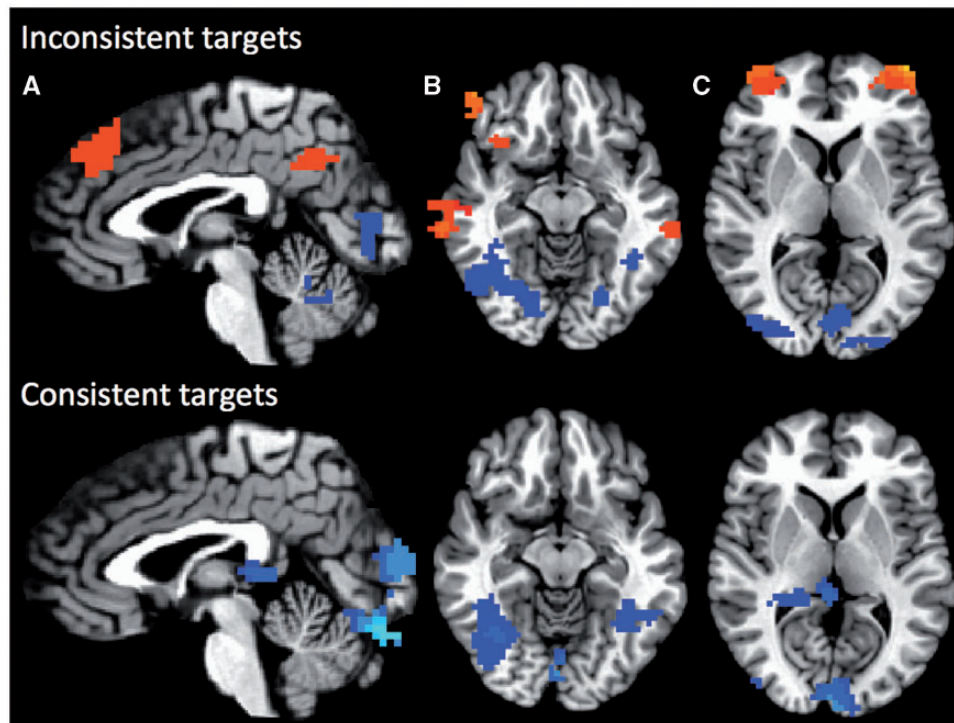


Fig. 3 Last two trials contrasted against first three trials, split by target type. Inconsistent targets displayed on top, consistent targets displayed on bottom. Hot activations represent stronger activation during the last two trials of each target, cold activations represent stronger activation during the first three trials of each target. Dorsomedial PFC, PCC/precuneus (A), anterior insula, bilateral STS (B), and bilateral rostrolateral PFC (C) all show stronger activity during the last two trials, compared to the first three trials, but only when participants were considering evaluatively inconsistent targets. Conversely, bilateral fusiform gyri showed stronger activity during the first three trials, compared to the last two trials, across both types of targets (B).

It seems likely that the STS and IPL are involved in processing specific to person targets in the context of this task. Conversely, the PCC and rLPFC are better suited to aid in more general, task-related processing during the updating impressions task.

While the PCC is typically associated with the default mode network (Gusnard and Raichle, 2001; Greicius *et al.*, 2003; Buckner *et al.*, 2008), it has also been implicated in a host of seemingly disparate processes, ranging from representation of subjective value (McCoy *et al.*, 2005; Kable and Glimcher, 2007; Levy *et al.*, 2010), to autobiographical memory retrieval (Maddock *et al.*, 2001), to goal-directed cognition (Spreng *et al.*, 2010). A recent reconceptualization of the PCC's function attempts to reconcile these various functions within one parsimonious explanation, suggesting that the PCC is critical for adapting to changes in the environment (Pearson *et al.*, 2011). This account of the PCC is extremely in step with the demands of the current experiment, wherein our participants had to identify relevant changes (i.e. behaviors inconsistent with existing impressions of person targets) and subsequently, adjust to those changes and act accordingly (i.e. update their impressions of person targets, as evidenced by changes in behavioral ratings).

The lateral PFC has also been linked to high-level cognitive processes, including maintaining abstract mental sets (Christoff *et al.*, 2007), multitasking (Burgess *et al.*, 2001; Burgess *et al.*, 2003; Badre *et al.*, 2004), and perhaps most importantly, the flexible exertion of cognitive control (Braver *et al.*, 2003; Braver *et al.*, 2009). Specifically, activity in the rostral portion of lateral PFC is associated with episodic control (Koechlin *et al.*, 2003; Kounieher *et al.*, 2009)—in which a previously encountered cue modifies the perception or interpretation of present stimuli (Egner, 2009). In the context of the present study, this conceptualization of rLPFC's role is particularly informative. The rLPFC activity in response to evaluatively inconsistent targets likely

reflects the influence of previously learned information on participants' evaluations of new information.

Limitations

Several low-level aspects of our design may be influencing our results. First and foremost, it is possible that the inclusion of trial-by-trial ratings is imposing an unnaturally high demand to update impressions upon our participants. While we concede that this is indeed a limitation of our approach, our intention was to collect a moment-to-moment measure of participants' impressions, so we could be absolutely certain that they showed behavioral evidence of updating. Future work could simply measure participants' impressions only once following the presentation of all five behaviors.

Second, we employed a control condition (faces presented alone) in which we do not account for the reading that participants have to do in the consistent and inconsistent conditions. We chose to perform the faces-plus-behaviors *vs* faces alone contrast because it is consistent with previous related work (Schiller *et al.*, 2009; Baron *et al.*, 2011). More importantly, while this confound is unavoidable for our fROI analysis, our whole-brain analyses do not depend on this contrast.

Convergence with recent work

As discussed previously, recent studies involving trait-inconsistent updating (Ma *et al.*, 2011) and category-inconsistent updating (Cloutier *et al.*, 2011b) bear a great deal of relevance for the present investigation. Comparing between these three studies, we note interesting convergence in the neuroimaging results, even though they focus on different types of inconsistency. As Figure 4 shows, all three studies observed higher dmPFC, IPL, STS, PCC and IPFC activity when targets were behaviorally inconsistent, compared to when they were consistent.

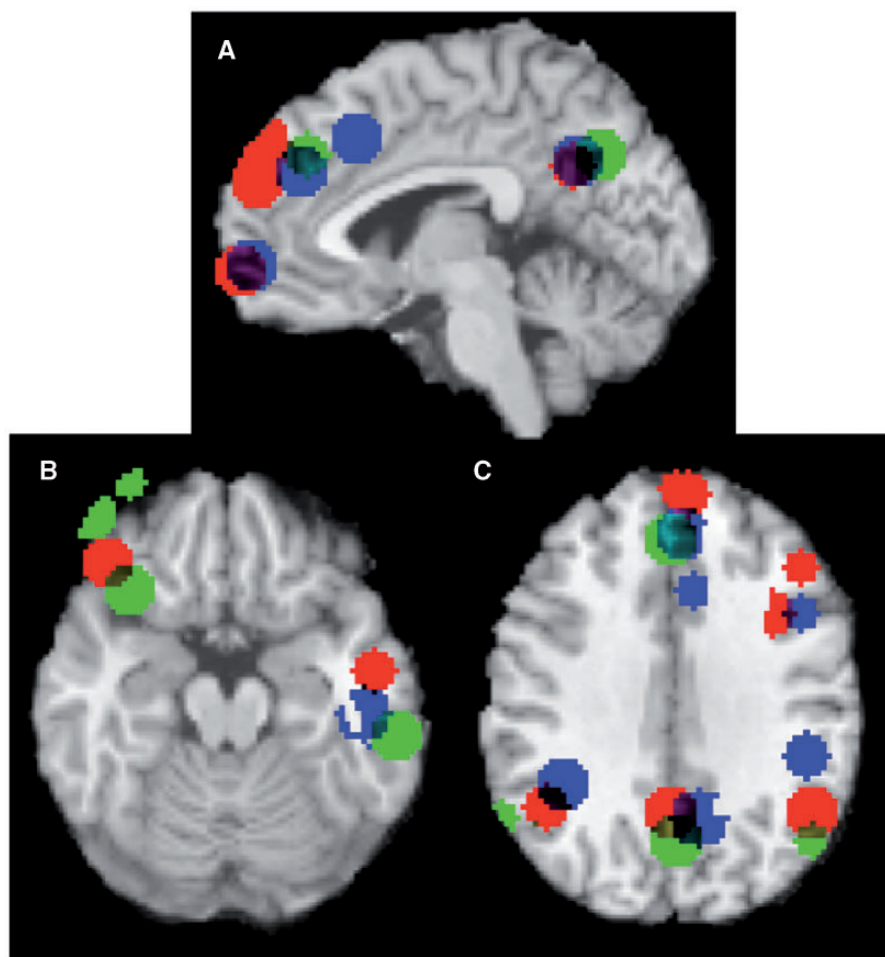


Fig. 4 Visualization of the overlap between three studies on impression updating—the present study; Ma *et al.* (2011); and Cloutier *et al.* (2011b). Peak voxels of each study were separately convolved with a 10 mm spherical kernel and subsequently overlaid on a canonical MRI image using meta-analytic software (Kober *et al.*, 2008). Note overlap in dmPFC, PCC/precuneus, mPFC (A), IPFC, STS (B) and IPL (C). Blue areas represent clusters reported by Ma and colleagues in the Trait Inconsistent > Trait Consistent (Intentional) contrast. Red areas represent clusters reported by Cloutier and colleagues in the Category Incongruent > Category Congruent contrast. Green areas represent clusters reported in the present study in the L2 > F3 (Inconsistent) contrast.

Previous work has observed additional inconsistency-related activity in a more posterior region of mPFC (referred to as domain-general pmFC; Ma *et al.*, 2011). One potential explanation for this divergence lies in the specific contrast with which Ma and colleagues obtained this result. While we chose to contrast the last two *vs* the first three trials in our behavior trajectories, they contrasted activity on only the critical fourth trial *between* target types (consistent *vs* inconsistent). In essence, the present analysis takes a more global perspective on the updating process as a whole, while Ma *et al.* (2011) isolated activity elicited at the precise moment when trait-inconsistent information was potentially presented. Running a similar analysis on our data yields activity in domain-general pmFC, as well (Supplementary Figure 4).

Taken together, these studies suggest that flexible updating of person impressions depends on the coordinated action of functional networks involved in social cognition and cognitive control. While this represents only a first step towards elucidating the neural dynamics underlying impression updating, a picture is beginning to come into focus, revealing a network of regions encompassing the dmPFC, IPL, STS, PCC and rPFC, associated with this process.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflict of Interest

None declared.

REFERENCES

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12, 169–77.
- Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4, 267–78.
- Badre, D., Wagner, A.D. (2004). Selection, integration, and conflict monitoring: assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron*, 41, 473–87.
- Baron, S.G., Gobbini, M.I., Engell, A.D., Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social, Cognitive, and Affective Neurosciences*, 6, 572–81.
- Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S. (2008). Associative learning of social value. *Nature*, 456, 245–9.
- Bliss-Moreau, E., Barrett, L.F., Wright, C.I. (2008). Individual differences in learning the affective value of others under minimal conditions. *Emotion*, 8, 479–93.
- Braver, T.S., Reynolds, J.R., Donaldson, D.I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, 39, 713–26.
- Braver, T.S., Paxton, J.L., Locke, H.S., Barch, D.M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences United States of America*, 106, 7351–6.
- Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1–38.

- Burgess, P.W., Quayle, A., Frith, C.D. (2001). Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia*, 39, 545–55.
- Burgess, P.W., Scott, S.K., Frith, C.D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: A lateral versus medial dissociation. *Neuropsychologia*, 41, 906–18.
- Calder, A.J., Young, A.W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews on Neuroscience*, 6, 641–51.
- Calder, A.J., Beaver, J.D., Winston, J.S., et al. (2007). Separate coding of different gaze directions in the superior temporal sulcus and inferior parietal lobule. *Current Biology*, 17, 20–5.
- Christoff, K., Keramian, K. (2007). Abstraction of mental representations: theoretical considerations and neuroscientific evidence. In: Bunge, S.A., Wallis, J.D., editors. *The Neuroscience of Rule-Guided Behavior*. New York: Oxford University Press.
- Cloutier, J., Kelley, W.M., Heatherton, T.F. (2011a). The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Social Neuroscience*, 6, 63–75.
- Cloutier, J., Gabrieli, J.D.E., O'Young, D., Ambady, N. (2011b). An fMRI study of violations of social expectations: When people are not who we expect them to be. *NeuroImage*, 57, 583–8.
- Cox, R. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29, 162–73.
- Decety, J., Chaminade, T., Grezes, J., Meltzoff, A.N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage*, 15, 265–72.
- Egner, T. (2009). Prefrontal cortex and cognitive control: motivating functional hierarchies. *Nature Neuroscience*, 12(7), 821–2.
- Engell, A.D., Haxby, J.V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, 45, 3234–41.
- Fiske, S.T., Linville, P.W. (1980). What does the schema concept buy us? *Personality and Social Psychology Bulletin*, 6, 543–57.
- Fiske, S.T. (1980). Attention and weight in person perception: the impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, 38, 889–906.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, 308, 662–7.
- Freeman, J.B., Schiller, D., Rule, N.O., Ambady, N. (2010). The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Human Brain Mapping*, 31, 150–9.
- Frith, U., Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B*, 365, 165–76.
- Fuhrman, R.W., Bodenhausen, G.V., Lichtenstein, M. (1989). On the trait implications of social behaviors: kindness, intelligence, goodness, and normality ratings for 400 behavior statements. *Behavior Research Methods, Instruments, & Computers*, 21, 587–97.
- Gallese, V., Keysers, C., Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8, 396–403.
- Ghazanfar, A.A., Chandrasekaran, C., Morrill, R.J. (2010). Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. *European Journal of Neuroscience*, 31, 1807–17.
- Gobbini, M.I., Haxby, J.V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45, 32–41.
- Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, 22, 1628–35.
- Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences USA*, 100, 253–8.
- Grossman, E.D., Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, 35, 1167–75.
- Gusnard, D.A., Raichle, M.E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews on Neuroscience*, 2, 685–94.
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4, 223–33.
- Hein, G., Knight, R.T. (2008). Superior temporal sulcus—it's my area: or is it? *Journal of Cognitive Neuroscience*, 20, 2125–36.
- Hoffman, E.A., Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 80–4.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H.L., Mazziotta, J.C., Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526–8.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., et al. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*. [Epub ahead of print; doi:10.1371/journal.pbio.0030079].
- Ishai, A. (2008). Let's face it: It's a cortical network. *NeuroImage*, 40, 415–9.
- Kable, J.W., Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10, 1625–33.
- Kanwisher, N., Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B*, 361, 2109–28.
- Kober, H., Barrett, L.F., Joseph, J., Bliss-Moreau, E., Lindquist, K., Wager, T.D. (2008). Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *NeuroImage*, 42, 998–1031.
- Koechlin, E., Ody, C., Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181–5.
- Kouneiher, F., Charron, S., Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 12, 939–45.
- Kayser, A. (1997). *Heads*. New York: Abbeville Press.
- Keysers, C., Wicker, B., Gazzola, V. (2004). A touching sight. SII/PV activation during the observation and experience of touch. *Neroun*, 42, 335–46.
- LaBar, K.S., Crupain, M.J., Voyvodic, J.T., McCarthy, G. (2003). Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*, 13, 1023–33.
- Leslie, K.R., Johnson-Frey, S.H., Grafton, S.T. (2004). Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage*, 21, 601–7.
- Levy, I., Snell, J., Nelson, A.J., et al. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology*, 103, 1036.
- Ma, N., Vanderkhorve, M., Baetens, K., Van Overwalle, F., Seurinck, R., Fias, W. (2011). Inconsistencies in spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*. [Epub ahead of print; doi:10.1093/scan/nsr064].
- Maddock, R.J., Garrett, A.S., Buonocore, A.H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience*, 104, 667–76.
- McCoy, A.N., Platt, M.L. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature Neuroscience*, 8, 1220–27.
- Mende-Siedlecki, P., Said, C., Todorov, A. (2011). The social evaluation of faces: a meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*. [Epub ahead of print; doi:10.1093/scan/nsr090].
- Miller, E.K., Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, 24, 4912–7.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2005). Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex. *NeuroImage*, 26, 251–7.
- Mitchell, J.P., Cloutier, J., Banaji, M.R., Macrae, C.N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience*, 1, 49–55.
- Montgomery, K.J., Haxby, J.V. (2008). Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 20, 1866–77.
- Pearson, J.M., Heilbronner, S.R., Barack, D.L., Hayden, B.Y., Platt, M.L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in Cognitive Sciences*, 15, 143–51.
- Pelphrey, K.A., Mitchell, T.V., McKeown, M.J., Goldstein, J., Allison, T., McCarthy, G. (2003a). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, 23, 6819–25.
- Pelphrey, K.A., Singerman, J.D., Allison, T., McCarthy, G. (2003b). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia*, 41, 156–70.
- Pelphrey, K.A., Morris, J.P., McCarthy, G. (2004a). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16, 1706–16.
- Pelphrey, K.A., Viola, R.J., McCarthy, G. (2004b). When strangers pass: processing of mutual and averted gaze in the superior temporal sulcus. *Psychological Science*, 15, 598–603.
- Pelphrey, K.A., Morris, J.P. (2006). Brain mechanisms for interpreting the actions of others from biological-motion cues. *Current Directions in Psychological Science*, 15, 136–40.
- Puce, A., Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of Royal Society of London, Series B, Biological Sciences*, 358, 435–45.
- Reeder, G.D., Brewer, M.B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, 86, 61–79.
- Reeder, G.D., Spores, J.M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, 44, 736–45.
- Ruby, P., Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, 4, 546–50.
- Ruby, P., Decety, J. (2003). What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *European Journal of Neuroscience*, 17, 2475–80.
- Rudoy, J.D., Paller, K.A. (2009). Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience*, 3, 1–6.
- Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, 10, 1–12.
- Saxe, R., Carey, S., Kanwisher, N. (2004). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 87–124.

- Schiller, D., Freeman, J.B., Mitchell, J.P., Uleman, J.S., Phelps, E.A. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, 12, 508–14.
- Skowronski, J.J., Carlston, D.E. (1987). Social judgment and social memory: the role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, 52, 689–99.
- Skowronski, J.J., Carlston, D.E. (1989). Negativity and extremity biases in impression formation: a review of explanations. *Psychological Bulletin*, 105, 131–42.
- Spreng, R.N., Stevens, W.D., Chamberlain, J.P. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, 53, 303–17.
- Srull, T.K., Wyer, R.S. (1989). Person memory and judgment. *Psychological Review*, 96, 58–83.
- Talairach, J., Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme.
- Todorov, A., Olson, I. (2008). Robust learning of affective trait associations with faces when the hippocampus is damaged, but not when the amygdala and temporal pole are damaged. *Social, Cognitive, and Affective Neuroscience*, 3, 195–203.
- Todorov, A., Uleman, J.S. (2002). Spontaneous trait inferences are bound to actors' faces: evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, 83, 1051–65.
- Todorov, A., Uleman, J.S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39, 549–62.
- Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45, 163–73.
- Uddin, L.Q., Molnar-Szakacs, I., Zaidel, E., Iacoboni, M. (2006). rTMS to the right inferior parietal lobule disrupts self-other discrimination. *Social, Cognitive, and Affective Neuroscience*, 1, 65–71.
- Vaina, L.M., Solomon, J., Chowdhury, S., Sinha, P., Belliveau, J.W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences United States of America*, 98, 11656–61.
- Verosky, S., Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, 21, 779–85.
- Wager, T.D., Nichols, T.E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage*, 18, 293–309.
- Wicker, B., Michel, F., Henaff, M.A., Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. *NeuroImage*, 8, 221–7.
- Zaki, J., Bolger, N., Weber, J., Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences United States of America*, 106, 11382–7.