Neural correlates of anchoring-and-adjustment during mentalizing

Diana I. Tamir[1] and Jason P. Mitchell

Department of Psychology, Harvard University, Cambridge, MA 02138

Recent studies have suggested that the medial prefrontal cortex (MPFC) contributes both to understanding the mental states of others and to introspecting about one's own mind. This finding has suggested that perceivers might use their own thoughts and feelings as a starting point for making inferences about others, consistent with "simulation" or "self-projection" views of social cognition. However, perceivers cannot simply assume that others think and feel exactly as they do; social cognition also must include processes that adjust for perceived differences between self and other. Recent cognitive work has suggested that such correction occurs through a process of "anchoring-and-adjustment" by which perceivers serially tune their inferences from an initial starting point based on their own introspections. Here, we used functional MRI to test two predictions derived from this anchoring-and-adjustment view. Participants (n = 64) used a Likert scale to judge the preferences of another person and to indicate their own preferences on the same items, allowing us to calculate the discrepancy between the participant's answers for self and other. Whole-brain parametric analyses identified a region in the MPFC in which activity was related linearly to this self–other discrepancy when inferring the mental states of others. These findings suggest both that the self serves as an important starting point from which to understand others and that perceivers customize such inferences by serially adjusting away from this anchor.

functional neuroimaging | self-reference | social cognition

Humans have a remarkable capacity for reading the minds of other humans. Despite limited access to the inner workings of others' mental states, we fluently infer others' thoughts, feelings, intentions, and personality traits and accomplish these inferential feats based on minimal information. Given how little direct access perceivers have to others' minds, a fundamental challenge in cognitive science is the understanding of how we successfully mentalize about what others are thinking and feeling. Several attempts to address this issue have started with the simple observation that, although we never enjoy immediate access to other minds, we do have extensive experience with a good proxy system: our own minds. As such, humans might draw on their own self-knowledge, experience, and mental states to understand the minds of others. Specifically, when predicting how someone might respond to a particular situation, perceivers might imagine their own thoughts and feelings in a mentally simulated version of that same experience and then assume that the other person would think and feel similarly.

That humans routinely use themselves as a source of information about others' minds is supported by a number of distinct empirical observations. Perceivers consistently assume that others hold the same opinions (1, 2), know the same facts (3), and engage in the same activities (4) as they themselves do (reviewed in ref. 5). Both children and adults have difficulty understanding others' thoughts and feelings when those mental states differ from their own (6, 7). Moreover, recent neuroimaging studies have revealed a common neural basis for introspecting about the self and for inferring the mental states of others (8). Specifically, ventral aspects of the medial prefrontal cortex (MPFC) respond preferentially both during self-referential processing (9, 10) and

during mental state inferences about others (11–13); studies that have directly compared self-referential thought and mentalizing have found considerable overlap between the neural correlates underlying both abilities (14–16). Together, these observations suggest that knowledge about oneself may serve as an important starting point from which to understand the minds of others.

Nevertheless, appropriate mental state inferences also must acknowledge the differences between self and others. Individuals are defined by their idiosyncratic beliefs and preferences (hence the old adage that one cannot argue with taste), and even two highly similar people will not hold identical opinions and attitudes. Accordingly, any account of how perceivers make use of their own mental states to understand those of others must address how one adjusts away from a starting point anchored on self. In other words, even if we start with our own thoughts, feelings, and opinions when contemplating those of others, we also typically need to correct for idiosyncratic aspects of that person's mind.

Empirical research has confirmed that individuals naturally engage in such correction. Despite a tendency to assume that others share their mental experiences, perceivers do not categorically conclude that others think and feel exactly the same way they themselves do. For example, when asked to estimate how others will vote, individuals tend to overestimate the number of peers who support the same candidate that they do but do not blithely believe that everyone else in the nation shares their political opinions (2). Perceivers also tailor their inferences about others on the basis of individuating information about them, for example, recognizing that men and women (17) or liberals and conservatives (14, 15) might hold different opinions both from each other and from the perceiver.

Recent cognitive work (3, 18) has suggested that perceivers generate individualized inferences about others through the process of "anchoring-and-adjustment," a mechanism first described in the context of nonsocial judgment and decision-making. When respondents are asked to make judgments with uncertain or indeterminate answers, they appear first to generate a plausible "anchor" value and then serially to adjust away from it (18, 19). For example, if asked to estimate the freezing point of vodka, an individual might first anchor on a relevant and well-known piece of information, 0 °C, the freezing point of water. However, because most people know that alcohol remains a liquid at lower temperatures than water, respondents will adjust their answer correspondingly downward. Tversky and Kahneman (19) further suggested that this adjustment mechanism will unfold serially, such that additional processing will be needed to reach values farther from an anchor point than those that more closely overlap with the point of origin.

Anchoring-and-adjustment also is thought to underlie inferences about mental states. Understanding the mind of another

PSYCHOLOGICAL AND COGNITIVE SCIENCES

NEUROSCIENCE

person shares a good deal in common with other kinds of (nonsocial) inferences. Mentalizing typically results in an uncertain inference: Perceivers can rarely know "for sure" what others are thinking or feeling but can generate plausible estimates of their mental states. Such inferences appear to originate from a self-generated anchor value that comprises one's own mental states. Because perceivers know that these anchor values do not overlap perfectly with the "correct" answer about what another person thinks or feels, they may need to adjust serially from their own mental states to customize their inferences about others.

Here, we use functional neuroimaging to examine these two latter predictions: (*i*) that perceivers use their own introspections as an anchor value from which to understand the mental states of others, and (*ii*) that perceivers adjust away from such anchor values serially, such that judgments far from the initial anchor will require greater cognitive processing than those close to the anchor. Across four independent data collections, participants were scanned while alternately reporting their own preferences and judging the preferences of another individual. Each item was judged once for self and once for another person, allowing us to measure the discrepancy between each participant's response for self and other on each item. This design allowed us to test both predictions simultaneously. First, to the extent that perceivers spontaneously use their own preferences and opinions as an anchor for inferring those of others, these self–other discrepancy values should capture meaningful variance in neural response. That is, the extent to which brain activity distinguishes at all between mental state inferences that differ from introspections about the self and those inferences that overlap with such introspections suggests that perceivers naturally consider their own mental states when mentalizing. On the other hand, if perceivers do not anchor on the self to mentalize about another person, then the discrepancy between responses for self and other should be psychologically (and, thus, neurally) meaningless. In other words, to the extent that mental state inferences draw (at least in part) on self-referential processing, we predicted that neural response during mentalizing about others would vary as a function of the discrepancy between self and other.

Second, to the extent that mentalizing entails serial adjustment away from one's own thoughts and feelings, we expected that any such relation between neural response and self–other discrepancy would be linear in nature. Because serial adjustment is thought to be a sequential process that occurs incrementally, we expect its involvement in mentalizing to be revealed by linear changes in neural response. To test these predictions, we treated the discrepancy between self and other as a parametric modulator in a whole-brain, random-effects analysis that identified any brain region in which activity during mentalizing about another person was predicted by the discrepancy in behavioral responses between other and self. In particular, to identify brain regions that plausibly might subserve the process of serially correcting an introspection about oneself into an individualized inference about another person, we looked for regions that demonstrated a linear relation of blood oxygenation level-dependent (BOLD) response to self–other discrepancy during mentalizing about others. We further confirmed that the response of such regions did not correlate with self–other discrepancy when introspecting about self.

On the basis of extant research, we hypothesized that anchoring and adjustment processes would be localized to the MPFC. As reviewed above, this region has been observed in nearly all studies that have described the neural basis of mentalizing (8). At the same time, this region contributes importantly to the ability to introspect about one's own mental states (9, 10, 20), and earlier studies have identified considerable overlap between the regions of the MPFC engaged by mentalizing about others and introspecting about self (14–16). Based on this earlier work, we hypothesized that this region also might subserve the processes that allow perceivers to individualize their inferences

about another person's mind by modifying an initial self-based anchor into a more appropriate, individually tailored inference.

Results

Behavioral Results. Cognitive research on the nature of anchoring-and-adjustment has suggested that adjustment proceeds serially over time and with some effort (18). Therefore, reaction time (RT) and cognitive effort present potential confounds in the current study: Highly discrepant inferences about others may require the exertion of cognitive effort over a longer period than inferences that overlap with self-based anchors. Belying this possibility, regression analysis with participant as a random variable demonstrated no significant relation between RT and discrepancy scores [$b = 0.000028$, $t(8235) = 1.59$, $P = 0.11$]. Likewise, no correlation was observed between RT and discrepancy across participants (mean $r = 0.005$). Together, these data rule out the possibility that any differences in neural activity as a function of self–other discrepancy may be caused by incidental differences in RT (but see below for extended discussion of the predicted effect of serial adjustment on RT).

Imaging Results. To identify brain regions in which BOLD response during judgments of others correlated with discrepancy from one's own self-reported opinion, we conducted a whole-brain random-effects analysis on the self–other discrepancy scores associated with Other trials. Consistent with our predictions, this analysis revealed extensive modulation of the MPFC, encompassing a region immediately anterior to the genu of the corpus callosum as well as a more dorsal subregion (Fig. 1). In other words, the further one's inference about another person was from one's own introspection about self, the greater was the activity of the MPFC. Self–other discrepancy also was correlated with activity in the right orbitofrontal cortex, left inferior frontal gyrus, and thalamus (Table 1). No brain regions showed the inverse pattern of decreasing activity with increasing discrepancy.

Our model suggests that, in attempting to understand the mental states of others, one first anchors on a self-based introspection and then adjusts from this point. However, it is logically possible that these cognitive events occur in reverse order, such that, in introspecting about one's own preferences, one anchors on and then adjusts from an inference about other people. To rule out this possibility, we also examined whether self–other discrepancy scores were related to neural activity during Self judgments with the expectation that such scores should be unrelated to BOLD response. From each of the regions identified above, we extracted parameter estimates reflecting both (*i*) self–other discrepancy during Other trials and (*ii*) self–other discrepancy during Self trials. Discrepancy during self-report was not correlated significantly with BOLD response in the MPFC, left inferior frontal gyrus, or thalamus (all $P > 0.20$) and correlated only marginally in the orbitofrontal cortex ($P = 0.09$). Moreover, self–other discrepancy during Other judgments was a significantly better predictor of activity than self–other discrepancy during Self judgments in both the MPFC and thalamus (both $P < 0.05$); this difference did not reach significance in inferior frontal gyrus ($P = 0.08$) and orbitofrontal cortex ($P = 0.15$; Table 1). Together, these results weigh against the possibility that the relation between the MPFC response and self–other discrepancy during inferences about others results from nonspecific, incidental factors correlated with discrepancy scores and confirm that these patterns of neural activity during judgments about others reflect adjustment from an anchor point derived from self-referential introspection rather than vice versa.

Finally, we conducted a whole-brain random-effects analysis to identify any neural region in which self–other discrepancy was correlated with BOLD response during Self trials. We observed a single, unexpected response in the region of the right inferior frontal gyrus [Montreal Neurological Institute (MNI)
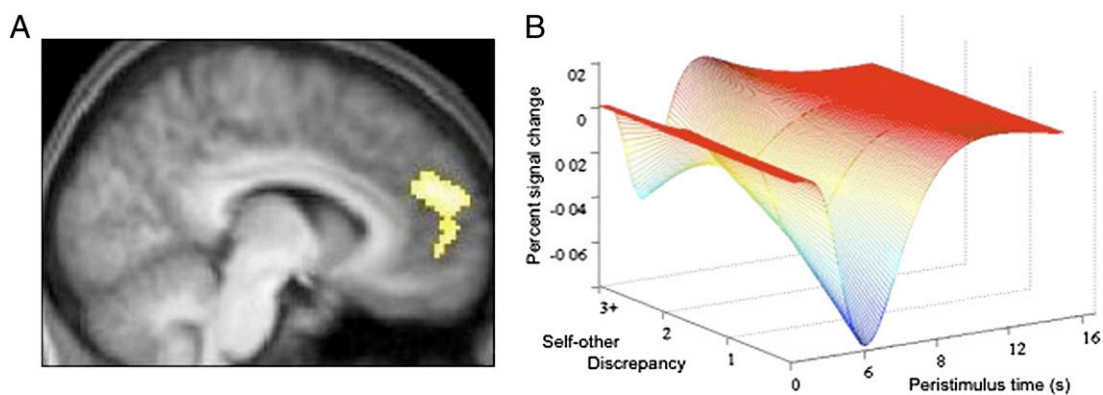
Fig. 1. Self–other discrepancy scores during Other trials were correlated with a sizeable region of the MPFC. A displays this region on a sagittal slice (*x* = 8) of participants' mean normalized brain. B displays the hemodynamic response of this region as a function of self–other discrepancy. The *x* and *y* axes of this graph represent estimated changes in BOLD response time-locked to trial onset. Consistent with most observations of MPFC response, signal change in this region took the form of deactivations relative to resting baseline (29). The *z* axis displays changes in the shape of this response across varying levels of self–other discrepancy.

coordinates: 50, 52, 27]. The relation between discrepancy scores and BOLD response in this region was marginally stronger for Self than Other trials [$t(63) = 1.95$, $P = 0.06$, $d = 0.25$], suggesting that, unlike the response of the MPFC, the response of this right inferior frontal region was more closely related to self–other discrepancy when introspecting about the self than when mentalizing about another person. Similar right frontal activation has been linked to a range of cognitive functions including semantic retrieval, stereotyping, categorization, response inhibition, assessment of emotional facial expressions, recognition of one's own face, humor appreciation, perception of vocal prosody, and thinking about one's own affective state (reviewed in ref. 21). The role of this region in introspecting about self vis-à-vis others is unknown at this time.

Because the MPFC cluster identified by the primary analysis included a considerable area of cortex (679 voxels, extending from $z = -12$ to $z = 35$), we defined separate regions of interest around local activation peaks (*Methods*). Three subregions were identified, two of which were centered relatively dorsally (MNI coordinates: 5, 50, 22 and 6, 60, 20, respectively) and one relatively ventrally (MNI: −10, 52, 8). We then examined the pattern of response in each of these regions of interest as a function of self–other discrepancy during Other trials. This analysis suggested a potential dissociation between dorsal and ventral MPFC. Whereas the two dorsal clusters showed a significantly linear pattern [$F(1, 62) = 5.83$, $P = 0.02$ and $F(1, 62) = 7.16$, $P = 0.01$, respectively], the ventral cluster demonstrated only a marginally significant linear correlation with self–other discrepancy [$F(1, 62) = 3.42$, $P = 0.07$]. (One participant was excluded from these analyses because his data did not include responses at all levels of

self–other discrepancy.) In contrast, the pattern of activity in the ventral cluster was best described by a thresholdlike response, whereby trials with zero discrepancy between self and other were significantly different from trials with any amount of discrepancy (1, 2, or 3+)—$F(1, 62) = 11.90$, $P = 0.001$—but discrepant trials were not significantly different from each other (all $P > 0.8$). The two patterns are illustrated in Fig. 2.

Discussion

The current study further refines the role of the MPFC in social cognition by showing that this region contributes to processes that subserve adjustment from self-based introspective anchors when inferring the mental states of others. In a series of data collections, participants alternately reported their own preferences and opinions and inferred the preferences and opinions of other people. Using the discrepancy between one's response to the same question for self and other as a predictor variable, we identified a small set of brain regions in which activity during inferences about others was related linearly to how different one judged the other person's preferences to be from one's own. In other words, the greater the activity in these regions when inferring the preferences and opinions of another person, the greater was the difference between the judgment about that person's and one's answer to the same question. Consistent with existing data showing its engagement during both introspections about self and mentalizing about others, the MPFC was the most sizeable and consistent region in this set (Table 1). The correlation between MPFC response and self–other discrepancy was observed even though the two judgments (Self and Other) took place in random order at different points in the experiment, sometimes separated by as much as 20 minutes.

Table 1. Regions significantly modulated by self–other discrepancy during Other judgments

Anatomic label	Axis			Volume	Maximum T	Other	Self (P)	$T_{Other > Self}$ (P)
	x	*y*	*z*					
Medial prefrontal cortex	6	60	20	679	4.56	2.40	0.62 (0.23)	2.56 (0.01)
Inferior frontal gyrus	−48	20	−14	139	4.43	2.68	0.80 (0.37)	1.80 (0.08)
Posterior orbital gyrus	28	22	−20	64	4.06	1.82	0.87 (0.09)	1.45 (0.15)
Thalamus	−8	−4	8	73	3.79	2.83	0.21 (0.82)	2.18 (0.03)

Coordinates refer to the Montreal Neurological Institute stereotaxic space. T-values reflect the degree to which BOLD activity during Other judgments is modulated by self–other discrepancy, as computed by SPM2. The three rightmost columns report, respectively: (*i*) parameter estimates associated with self–other discrepancy on Other trials, (*ii*) parameter estimates associated with self–other discrepancy on Self trials, and (*iii*) the results of a *t* test comparing the two. For example, in medial prefrontal cortex, self–other discrepancy on Self trials was not significantly different from zero ($P = 0.23$) and was significantly higher on Other trials than on Self trials ($P < 0.01$).
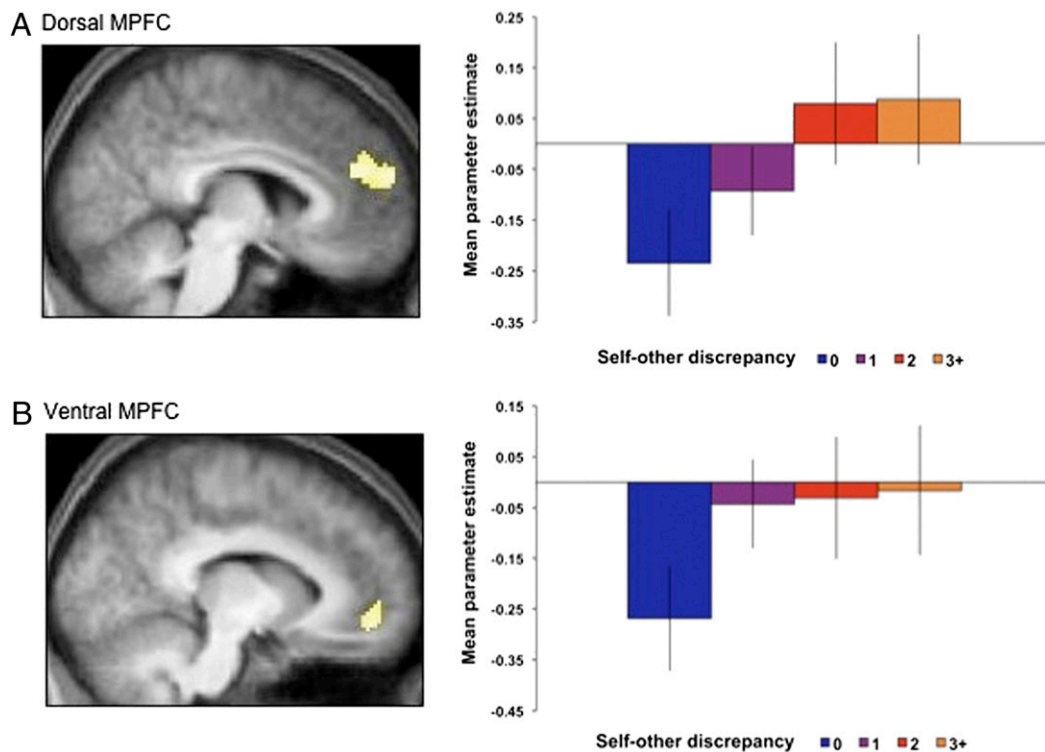
Fig. 2. The relation between BOLD response and self–other discrepancy during Other trials was calculated separately for subregions of the MPFC. Although the response of dorsal MPFC (*A*) increased linearly with increasing self–other discrepancy, the response of ventral MPFC (*B*) distinguished only between trials on which self–other discrepancy was zero (overlap between self and other) versus greater than zero (discrepancy between self and other). Error bars indicate the SEM.

Secondary analyses demonstrated that this effect was not attributable to nonspecific features of self–other discrepancy observable for Self trials and did not correlate with differences in cognitive effort as indexed by RT.

These results support two claims about the nature of social cognition. First, the observation that neural response when mentalizing about another person is related directly to the discrepancy from one's own answers strongly suggests that self-referential introspection contributes to social cognition. If perceivers did not consider their own preferences spontaneously when mentalizing about those of others, it is unclear why any neural activity should correlate with self–other discrepancy. As such, these data augment the growing number of cognitive and neuroimaging studies that suggest that one important strategy for understanding the mind of another person is by reference to one's own simulated thoughts and feeling (14–16). It is unclear how other possible strategies for inferring the mental states of others, such as theorylike reasoning (22, 23), could account for this finding.

Second, these results suggest a role for the MPFC in transforming self-based introspections into an individualized inference about another person. As a cognitive process underlying judgment and decision-making generally, "adjustment" is thought to take place in a serial fashion, such that respondents must engage in effortful cognitive processing for each "step" away from the anchor point. The current data support the hypothesis that mentalizing likewise relies on serial adjustment: The farther a perceiver adjusted from an anchor based on the perceiver's own mental states, the more processing, as indexed by MPFC response, was observed.

Because serial adjustment is defined as a process that occurs incrementally over time, a number of earlier studies have used RT to document the presence of serial adjustment; in general, adjustment correlates with the length of time that a participant takes to make a response. However, we observed no such correlation between self–other discrepancy and RT in the current

study. In all likelihood, this null finding reflects the constraints of the event-related functional MRI (fMRI) designs used in these studies. Participants were obliged to read and respond within a relatively short window (typically, 2,650 ms), thus compressing RT variability and undermining the ability to detect correlations between RT and discrepancy scores. Consistent with this speculation, we recently documented a significant relation between RT and self–other discrepancy in a series of cognitive studies in which participants were given a considerably longer response window (24). Under such conditions, we have observed greater self–other discrepancy reliably associated with longer RTs. Nonetheless, the observation that self–other discrepancy was independent of RT in the current studies rules out the possibility that greater neural response during highly discrepant trials merely indicates greater "time on task."

Secondary analyses revealed that the linear effect of self–other discrepancy on neural response was associated most closely with relatively dorsal aspects of the MPFC. Specifically, whereas the dorsal MPFC demonstrated a strong linear (i.e., serial) relation with self–other discrepancy, more ventral aspects of the MPFC instead showed a thresholdlike response, with the lowest activity for discrepancy scores of zero and equivalently higher activity for any score greater than zero. Although somewhat provisional, these results comport with earlier suggestions of a "division of labor" within the MPFC, such that relatively dorsal regions subserve mentalizing about others perceived to be particularly dissimilar from self (15, 16). In the current study, the response of dorsal MPFC correlated with the degree to which a target individual was perceived to be dissimilar from self on a particular item. That is, earlier work had established that the dorsal MPFC was engaged preferentially when thinking about a person who was consistently viewed as dissimilar from the self, and the current data extend these findings to suggest that similar regions may be sensitive to dissimilarity on an item-by-item basis (e.g., distin-

Tamir and Mitchell

guishing between opinions that a target person shares with the self versus those that diverge from the self). Interestingly, earlier findings tended to locate mentalizing about dissimilar others in subregions of the MPFC that are even more dorsal than those observed in the current study, suggesting a potential dissociation between perceiving another person to be uniformly dissimilar from the self and perceiving another person to be dissimilar from self only on specific dimensions. For example, one may perceive another person as generally similar to self and yet infer that that person happens not to share one's love of snowboarding (or, inversely, perceive someone as dissimilar but infer that that person does think about a particular issue in the same way as oneself). The region of dorsal MPFC observed here may subserve judgments about others that happen to differ along a particular dimension rather than judgments about globally dissimilar others. By comparison, even more dorsal regions of MPFC may participate in understanding others in a manner that does not anchor on self knowledge when such individuals are globally dissimilar from self. In either event, the particular dorsal MPFC region observed here appears to contribute to mentalizing in a manner anchored on the self (15).

The threshold response evidenced by the ventral MPFC was less consistent with a role for this region in serial adjustment. We speculate that this region may signal specifically the need for responses that are equivalent to self (discrepancy values of zero) rather than discrepant from self (i.e., discrepancy values greater than zero), but the current data do not offer a specific model of how this signaling is accomplished. Unexpectedly, the pattern of responses observed in the ventral MPFC appear to contradict earlier findings that this region is maximally engaged when considering a like-minded other whose opinions overlap with one's own (15). Again, we suggest that there may be important differences between contemplating a person viewed as globally similar to (or dissimilar from) the self, in contrast to considering specific domains with which another person's opinion overlaps with or diverges from one's own. However, these putative differences between dorsal and ventral MPFC are fairly speculative, and future research will be necessary to investigate possible differences in how each of these regions contributes to self-based mentalizing.

Nevertheless, the current findings shed light on the cognitive processes brought to bear when considering other minds. Fifteen years ago, researchers first established that a relatively small number of brain regions—the MPFC, temporo-parietal junction, and precuneus—were preferentially engaged during mentalizing (25, 26). More recently, researchers have begun to specify exactly how these regions give rise to social cognition; among the most promising possibilities is that the MPFC subserves the use of self-projective simulation as one route to understanding other minds (27). Here, we push these observations one step forward by suggesting that subregions of the MPFC not only may use the self as an anchor point from which to understand others but also may actively allow perceivers to adjust their inferences about another person. As such, these data suggest a solution to one problem of any simulation account of social cognition: How do perceivers avoid the simple assumption that others will think and feel exactly as they do? Here, we demonstrate that—as anticipated earlier by social psychologists (3, 18)—this challenge may be met through a process of anchoring-and-adjustment similar to that described for nonsocial inferences (19) and subserved by regions of the MPFC important for social cognition more generally.

Methods

Participants. Sixty-four individuals (38 female) participated in one of four separate data collections, two of which have been published earlier (15, 17). Participants were all right-handed, native English speakers with no history of neurological impairment and normal or corrected to normal vision (mean age, 21.2 y; range, 18–26 y). Informed consent was obtained from all participants in a manner approved by the Human Studies Committee of the Mas-

sachusetts General Hospital or the Committee on the Use of Human Subjects at Harvard University.

Design. Although the specific design of the four experiments differed slightly, each required participants to answer a series of questions about their opinions and preferences and to judge how other individuals would answer the same questions (Table S1). On each trial, participants saw a cue that indicated the target of the judgment (self or another person) and a brief phrase (e.g., "enjoy winter sports such as skiing or snowboarding"; "fear speaking in public"). Participants used either a four- or five-point scale either to report how well the statement described themselves or to judge how well it described the other person. Within each experiment, participants considered the same set of statements for self and other. Randomization ensured that each statement had an equal chance of being answered first for self and then subsequently for another person or first for other and then for self (because no differences in presentation order were detected, results are collapsed across this factor). The duration of each trial was 3,800 ms, and trials were separated by a variable interstimulus interval between 200–10,000 ms (28).

Before scanning, participants were told that the purpose of the experiment was to examine how people make inferences about target individuals on the basis of minimal or no information. In all studies, targets were college-aged individuals depicted by a photograph downloaded from an internet dating website, although the specific identity of individuals varied across studies. Across studies, target individuals were white and matched the gender of the participant, with three exceptions: In one study half the targets were Asian, in another study all participants judged both male and female targets, and in a third study all the targets were male. In two studies, participants were introduced to the target individuals through a short descriptive paragraph about their sociopolitical orientation; no information about the targets was provided in the other two studies. Thus, the wide range of specific target individuals and experimental conditions used across studies ensures the generalizability of findings, although this methodological variability also represents a potential source of statistical "noise" that may increase the possibility of type II errors (i.e., false negatives).

Imaging Procedure. In all studies, functional data were acquired using a gradient-echo echo-planar pulse sequence (TR = 2 s; TE = 35 ms) on either 1.5-T Siemens Sonata (one study) or 3-T Siemens Trio scanner (three studies). Images were acquired using either 26 or 31 axial, interleaved slices with a thickness of 5 mm (1-mm skip) and 3.75×3.75 mm or 3×3 mm in-plane resolution. Functional images were preprocessed and analyzed using SPM2 (Wellcome Department of Cognitive Neurology, London). Data first were preprocessed to correct for differences in slice-time acquisition in each whole-brain volume and were realigned spatially to correct for head movement. Images were then normalized to a standard anatomical space (3-mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). Normalized images then were smoothed spatially using an 8-mm FWHM Gaussian kernel.

Preprocessed images were analyzed using a general linear model in which the event-related designs were modeled using a canonical hemodynamic response function, its temporal derivative, and covariates of no interest (session mean and linear trend). Primary analyses identified voxels in which BOLD response was related linearly to the discrepancy between judgments of self and of others. Specifically, each judgment of another person (Other trial) was compared with the participant's own self-reported response to the same statement (Self trial). The absolute difference between Self and Other responses served as the measure of interest; for example, if a participant indicated that her fear of public speaking was a 4 but judged the other person would not fear it at all (1), then the discrepancy score for that trial was 3. To accommodate the different response scales used across experiments, discrepancy scores in each experiment were scaled such that the minimum and maximum values were 0 and 3, respectively. Trials on which a participant made no response to a particular statement for either Self or Other were excluded from analysis (9% of trials).

Discrepancy scores were modeled as parametric modulators for Other trials, and analyses identified brain regions in which activity was a linear function of discrepancy from Self. For comparison purposes, discrepancy scores also were included in analyses as modulators for Self trials; this analysis enabled us to identify any unexpected regions in which BOLD response during self-referential introspection was informed by judgments about another person. These analyses were performed individually for each participant, and the resulting contrast images were entered subsequently into a second-level analysis, treating participant as a random effect. Peak coordinates were identified using a statistical criterion of 50 or more contiguous voxels at a voxelwise threshold of $P < 0.001$, resulting in a corrected

experimentwise alpha level of $P < 0.05$. Functional regions of interest were defined using an automated search algorithm (R. Poldrack, University of California, Los Angeles) that identified peak activations separated by a minimum of 8 mm.

1. Krueger J, Clement RW (1994) The truly false consensus effect: An ineradicable and egocentric bias in social perception. *J Pers Soc Psychol* 67:596–610.
2. Ross L, Greene D, House P (1977) The false consensus effect: An egocentric bias in social perception and attribution processes. *J Exp Soc Psychol* 13:279–301.
3. Epley N, Keysar B, Van Boven L, Gilovich T (2004) Perspective taking as egocentric anchoring and adjustment. *J Pers Soc Psychol* 87:327–339.
4. Allport FH (1924) *Social Psychology* (Houghton Mifflin, Cambridge, MA).
5. Nickerson RS (1999) How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychol Bull* 125:737–759.
6. Wellman HM, Cross D, Watson J (2001) Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev* 72:655–684.
7. Niedenthal PM, Halberstadt JB, Margolin J, Innes-Ker ÅH (2000) Emotional state and the detection of change in facial expression of emotion. *Eur J Soc Psychol* 30:211–222.
8. Mitchell JP (2009) Social psychology as a natural kind. *Trends Cogn Sci* 13:246–251.
9. Kelley WM, et al. (2002) Finding the self? An event-related fMRI study. *J Cogn Neurosci* 14:785–794.
10. Northoff G, et al. (2006) Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage* 31:440–457.
11. Frith CD, Frith U (1999) Interacting minds—a biological basis. *Science* 286:1692–1695.
12. Gallagher HL, et al. (2000) Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21.
13. Gallagher HL, Jack AI, Roepstorff A, Frith CD (2002) Imaging the intentional stance in a competitive game. *Neuroimage* 16:814–821.
14. Jenkins AC, Macrae CN, Mitchell JP (2008) Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci USA* 105:4507–4512.
15. Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50:655–663.
16. Mitchell JP, Banaji MR, Macrae CN (2005) The link between social cognition and self-referential thought in the medial prefrontal cortex. *J Cogn Neurosci* 17:1306–1315.
17. Mitchell JP, Ames DL, Jenkins AC, Banaji MR (2009) Neural correlates of stereotype application. *J Cogn Neurosci* 21:594–604.
18. Epley N, Gilovich T (2001) Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychol Sci* 12:391–396.
19. Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–1131.
20. Fellows LK, Farah MJ (2007) The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? *Cereb Cortex* 17:2669–2674.
21. Mitchell JP, Ames DL, Jenkins AC, Banaji MR (2009) Neural correlates of stereotype application. *J Cogn Neurosci* 21(3):594–604.
22. Gopnik A, Wellman H (1992) Why the child's theory of mind is really a theory. *Mind and Language* 7:145–171.
23. Saxe R (2005) Against simulation: The argument from error. *Trends Cogn Sci* 9:174–179.
24. Tamir DI, Mitchell JP (2009) Anchoring and adjustment in mental state inference. *Annual Meeting of the Cognitive Science Society*, July 29–August 1, 2009, Amsterdam, The Netherlands.
25. Fletcher PC, et al. (1995) Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition* 57:109–128.
26. Goel V, Grafman J, Sadato N, Hallett M (1995) Modeling other minds. *Neuroreport* 6:1741–1746.
27. Mitchell JP (2009) Inferences about mental states. *Philos Trans R Soc Lond B Biol Sci* 364:1309–1316.
28. Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114.
29. Gusnard DA, Raichle ME, Raichle ME (2001) Searching for a baseline: Functional imaging and the resting human brain. *Nat Rev Neurosci* 2:685–694.