Neural Correlates of Stereotype Application

**Jason P. Mitchell, Daniel L. Ames, Adrianna C. Jenkins,
and Mahzarin R. Banaji**

Abstract

■ Recent research has focused on the disparate mechanisms that support the human ability to ''mentalize'' about the thoughts and feelings of others. One such process may rely on precompiled, semantic beliefs about the characteristics common to members of a social group, that is, on *stereotypes*; for example, judging that a woman may be more likely than a man to have certain interests or opinions. In the current study, we identified a pattern of neural activity associated with the use of stereotypes to judge another person's psychological characteristics. During fMRI scanning, participants mentalized about the likely responses of a female and male target to a series of questions, some of which were related to gender stereotypes (e.g., ''enjoys shopping for new clothes''). Trials on which participants applied a stereotype were segregated from those on which participants avoided stereotype use. The BOLD response in an extensive region of the right frontal cortex differentiated stereotype-applied from -unapplied trials. Moreover, this neural difference was correlated with a behavioral index of gender associations—the Implicit Association Test—administered after scanning. Results suggest that stereotype application may draw on cognitive processes that more generally subserve semantic knowledge about categories. ■

INTRODUCTION

Successful human social interaction relies on the ability to infer the hidden mental states of others (Dennett, 1987), that is, to *mentalize* about the thoughts, feelings, goals, desires, and preferences of other people. Recently, researchers have suggested that when mentalizing about others' minds, perceivers frequently use their own thoughts and feelings as a basis for understanding those of others, a process known as ''simulation'' or ''projection'' (Gallese, 2007; Davies & Stone, 1995). For example, perceivers tend both to assume that others know the same things they do (Epley, Keysar, Van Boven, & Gilovich, 2004; Nickerson, 1999) and to judge the emotional states of others in line with their own current moods (Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001; Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000). Moreover, recent neuroimaging research has suggested that overlapping patterns of neural activation accompany mentalizing about others and introspecting about one's own mental states (Mitchell, Macrae, & Banaji, 2006; Mitchell, Banaji, & Macrae, 2005).

However, simulation need not be the only route to understanding the mind of another person (Mitchell, 2005; Saxe, 2005). Perceivers may also base their judgments of others on consensual social knowledge about the groups to which an individual belongs, that is, on a *stereotype*. For example, I might not have sufficient experience with my friend Becky to know her taste in movies, but if I believe that, as a group, women generally dislike car chases and violence, I may rely on that information to infer that she would not fully appreciate seeing the latest summer blockbuster. Although stereotypes may prove an erroneous guide to any particular person's mind (perhaps Becky actually loves action flicks), psychologists have spent decades documenting the regularity with which perceivers judge others on the basis of their category membership (Greenwald & Banaji, 1995; Hamilton, 1981; Allport, 1954; Katz & Braly, 1933). Much of this research has suggested that perceivers frequently use stereotypes as heuristic shortcuts for understanding others' behavior or psychological characteristics, in lieu of the more complex individuating processes needed to fully consider the idiosyncrasies of another mind (Macrae, Bodenhausen, & Milne, 1994; Fiske & Neuberg, 1990).

As such, emerging accounts of the cognitive basis of mentalizing have produced something of a hybrid view of stereotyping. On one hand, many stereotypes refer to the putative mental aspects of groups of other people (e.g., their personalities and preferences) and likely form an important part of our cognitive repertoire for mentalizing about others' minds. On the other hand, stereotypes can be thought of as a form of precompiled, semantic knowledge about the world. Just as one may know that elephants are more likely to be found in Africa than in Antarctica, one may likewise ''know'' that men are more likely to be found at action movies and football games than at tearjerkers and shoe stores. What, then, are the cognitive processes that give rise to stereotype-based

Harvard University

© 2008 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 21:3, pp. 594–604

judgments of others? Specifically, do stereotypes about others' minds rely on the same kinds of social–cognitive processes as other routes to mentalizing? Or do they more closely resemble other forms of general, semanticized knowledge, as consistently suggested by research in social psychology?

This question can be adjudicated using the well-established pattern of neural activation associated with mentalizing. Across a sizeable number of studies, tasks that have required participants to consider the mental states of others have consistently been associated with activation in a set of brain regions that includes the medial prefrontal cortex, the temporo-parietal junction, and the precuneus/posterior cingulate cortex (Buckner, Andrews-Hanna, & Schacter, 2008; Gallagher & Frith, 2003; Saxe & Kanwisher, 2003; Fletcher et al., 1995; Goel, Grafman, Sadato, & Hallett, 1995), as well as the superior temporal sulcus and the temporal poles. In contrast, tasks that orient perceivers to the nonsocial aspects of a situation typically fail to engage these regions, instead activating those more closely associated with perception, attention, and working memory. For example, in a study of social versus nonsocial semantic knowledge, Mitchell, Heatherton, and Macrae (2002) compared the pattern of brain activity associated with judgments of people (specifically, words that could potentially describe the mental states of another person) with that associated with judgments of inanimate objects (fruit and clothing). Whereas judgments regarding the mental states of other people modulated both the medial prefrontal cortex and the temporo-parietal junction, judgments of inanimate objects were associated with activations in regions typically observed during semantic memory tasks, including lateral prefrontal (Maril, Wagner, & Schacter, 2001; Cabeza & Nyberg, 2000; Gabrieli, Poldrack, & Desmond, 1998; Wagner et al., 1998) and inferotemporal (Haxby et al., 2001; Chao, Haxby, & Martin, 1999; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999) cortices.

In the current study, we follow up on this observation by examining whether stereotypes function more similarly to other forms of mentalizing or to other forms of semantic knowledge. During fMRI scanning, participants considered the preferences and opinions of two targets: one man and one woman. On each trial, participants were asked to judge how likely the target would be to agree with a series of opinion questions, each of which was pretested to be either gender-stereotypical (e.g., "likes scented candles?"; "enjoys watching football?") or equally applicable to both sexes (e.g., "likes going to concerts?"). Importantly, this design allowed us to examine not merely those situations in which perceivers *could* judge another person on the basis of a stereotype, but those situations in which they actually *did* apply such a stereotype. Specifically, participants' behavioral responses were used to segregate trials on which they applied a stereotype (e.g., judging that a male target

would enjoy a stereotypically masculine activity or that he would dislike stereotypically feminine one) from those on which they judged the target in a manner inconsistent with gender stereotypes (e.g., judging that a female target would enjoy a stereotypically masculine activity, such as watching an action movie).

Lastly, because perceivers vary in the strength of their stereotypic associations, we included a series of postscanning behavioral measures to index how strongly each participant naturally associated men and women with distinct skills and proclivities (career vs. home; science vs. the humanities). We reasoned that those participants with especially strong gender-stereotypical associations would be those most likely to reveal neural differences associated with the application of stereotypes. Just as experts in nonsocial domains such as entomology or sports maintain especially rich semantic systems dedicated to their specialized knowledge, we hypothesized that perceivers who naturally use gender to divide the social world would have especially pronounced differences associated with stereotype-consistent vs. -inconsistent responses. Results demonstrated that, consistent with the view of stereotypes as semantic knowledge about social categories, activity in an extensive area of the right lateral frontal cortex not only distinguished the application of stereotypes from stereotype-inconsistent responses but also correlated with the strength of participants' gender associations, as measured postscanning.

METHODS

Participants

Participants were 17 (11 women) right-handed, native English speakers with no history of neurological problems (mean age = 20.1 years, range = 18–23 years). Informed consent was obtained in a manner approved by the Human Studies Committee of the Massachusetts General Hospital.

Stimuli

The stimulus set comprised 160 statements describing common traits, attitudes, and preferences. Stereotypical items ($n = 80$) referred to domains in which men and women are commonly believed to hold disparate opinions from each other (e.g., "like shopping for clothes"; "like action movies"). Nonstereotypical items ($n = 80$) referred to domains in which no such gender differences are acknowledged (e.g., "enjoy drinking coffee in the morning"; "like Coke better than Pepsi"). Stereotypical and nonstereotypical items were identified through pilot testing, in which a separate group of participants ($n = 20$) rated the gender stereotypicality of each statement on a 7-point scale, anchored by 1 = extremely masculine, 4 = neutral, and 7 = extremely feminine. For each item, pilot participants were instructed to indicate

the extent to which the statement conveyed a widely held gender stereotype in society at large, rather than the extent to which they personally believed the statement to be true of men and women. Items with a mean rating between 1 and 3 were designated *stereotypically masculine* ($M = 2.12$). Items with a mean rating between 3 and 5 were designated *nonstereotypical* ($M = 4.01$). Items with a mean rating between 5 and 7 were designated *stereotypically feminine* ($M = 5.95$). The full set of items is reproduced as Supplementary Material.

Behavioral Procedure

During scanning, participants performed a modified version of the opinion-judging task used by Mitchell et al. (2006). Each trial consisted of a photograph of either a man or a woman, presented above an opinion question. Photographs were simple black-and-white headshots of college-aged Caucasian individuals; two different faces were selected randomly for each participant from a pool of female and male faces. Participants were asked to use a 4-point scale (1 = least and 4 = most) to estimate how likely the target would be to endorse the opinion. The target and opinion question remained onscreen together for 3450 msec, during which time participants were obliged to make their response.

On an additional half of trials, participants reported their own opinions on the same set of questions. These self-referential trials were cued with a chalk outline of a head with the word ''me'' written inside used to represent the participant herself or himself. Such trials served to minimize the possibility that participants would spontaneously realize that the experiment was designed to examine gender stereotyping, and were not analyzed further. Participants answered all 160 questions for self and 80 questions each for the female and male targets (40 nonstereotypical, 20 stereotypically feminine, 20 stereotypically masculine). To optimize estimation of the event-related fMRI response, all trial types were intermixed in a pseudorandom order and separated by a variable interstimulus interval (450–9550 msec) (Dale, 1999), during which participants passively viewed a fixation crosshair.

To increase involvement in the task, participants were told that we knew the targets' actual responses to each of the opinion questions and that, following earlier research on the accuracy of first impressions, we were interested in examining the neural processes associated with accurate and inaccurate interpersonal judgments. In actual fact, all stimulus materials were newly created for this experiment. At no point before or during scanning did we mention the term stereotype or suggest that half the opinion questions referred to typically feminine or masculine preferences.

After scanning, participants completed two versions of the Implicit Association Test (IAT; Greenwald, McGhee,

& Schwartz, 1998), designed to measure the strength of each participant's automatic gender associations. The IAT assesses the conceptual association between two classes of stimuli by measuring differences in the speed with which participants can make the same behavioral response to exemplars from two categories (e.g., pressing the same button for pictures of snakes and positively valenced words compared to pressing the same button for snakes and negatively valenced words). On the *career/family* IAT, participants categorized an exemplar from one of four categories of stimuli: typically female names (e.g., Emily, Michelle), typically male names (e.g., Ben, Jeffrey), words that denoted family (e.g., home, children), or words that denoted career (e.g., office, business). In a block of stereotype-consistent trials, female names required the same behavioral response as family words (''d'' key), whereas male names required the same response as career words (''k'' key). In a block of stereotype-inconsistent trials, female names required the same response as career words, whereas male names required the same response as family words. The *science/arts* IAT was designed identically, except that words denoting science (e.g., astronomy, chemistry) or the humanities (e.g., history, arts) were used instead of those denoting career/family, and words denoting males and females (e.g., man, uncle; woman, daughter) were used instead of proper names. Each IAT block comprised 60 trials (15 each of the four trial types), and each participant completed the IAT blocks in a different random order. For both IATs, the strength of a participant's gender associations was indexed as the mean response latency to trials in the stereotype-inconsistent block minus the mean response latency to trials in the stereotype-consistent block. As such, higher IAT difference scores indicate a stronger association between female and family/humanities and between male and career/science. Because of the close conceptual similarity between the two IAT tasks, we averaged each participant's difference score on both into a single composite index of gender stereotypicality.

Finally, participants responded to five explicit self-report questions that asked them to judge how similar they were to (i) the typical member of their own sex; (ii) the typical member of the other sex; (iii) the male target used in the experiment; (iv) the female target used in the experiment; and one additional question that asked about (v) the importance of gender to their identity. For each question, participants indicated their answer on a 5-point scale, on which 1 represented the lowest (e.g., least similar) and 5 the highest (e.g., most similar) response.

Imaging Procedure

fMRI data were collected using a 3-Tesla Siemens Trio scanner. The task comprised 4 functional runs of 210 volume acquisitions (26 axial slices, 5 mm thick; 1 mm

skip). Functional imaging used a gradient-echo, echo-planar pulse sequence (TR = 2 s; TE = 35 msec; 3.75 × 3.75 in-plane resolution). Following the functional scans, we collected a high-resolution T1-weighted structural scan (MP-RAGE). PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) for Mac OS X (L. Bonatti, International School of Advanced Studies, Trieste, Italy) was used to project stimuli onto a screen at the end of the magnet bore, which participants viewed via a mirror mounted on the head coil. A pillow and foam cushions were placed inside the coil to minimize head movement.

fMRI data were preprocessed and analyzed using SPM2 (Wellcome Department of Cognitive Neurology, London, UK). First, functional data were time-corrected for differences in acquisition time between slices for each whole-brain volume and realigned to correct for head movement. Functional data were then transformed into a standard anatomical space (3-mm isotropic voxels) on the basis of the ICBM 152 brain template (Montreal Neurological Institute). Normalized data were then spatially smoothed (8 mm FWHM) using a Gaussian kernel.

Statistical analyses were performed using the general linear model in which the event-related design was modeled using a canonical hemodynamic response function, its temporal derivative, and additional covariates of no interest (a session mean and a linear trend). This analysis was performed individually for each participant, and contrast images for each participant were subsequently entered into a second-level analysis treating participants as a random effect. Peak coordinates were identified using a statistical criterion of 25 or more contiguous voxels at a voxelwise threshold of $p < .001$. This cluster size was selected on the basis of a Monte Carlo simulation of our brain volume (S. Slotnick, Boston College), which indicated that this cluster extent cutoff provided an experiment-wise threshold of $p < .05$, corrected for multiple comparisons.

Trials were conditionalized as a function of whether participants had the opportunity to stereotype another person and, if so, whether they applied the stereotype. Specifically, for judgments made about other people, stereotypical trials were segregated into those trials on which a participant judged a target to have an opinion consistent with her or his sex (stereotype-applied trials) versus an opinion inconsistent with her or his sex (stereotype-unapplied). For example, a stereotypically feminine item (e.g., ''enjoys scented candles'') was considered stereotype-applied if a participant judged that the female target was likely to agree (a response of ''3'' or ''4'') or that the male target was unlikely to agree with the question (a response of ''1'' or ''2''). In the same way, a stereotypically masculine item was considered stereotype-applied if a participant judged that the male target was likely or that the female target was unlikely to agree with the question. Conversely, a stereotypically feminine item was considered stereotype-unapplied if a

participant judged that the female target was unlikely to agree or that the male target was likely to agree with the question (and vice versa for stereotypically masculine items). The conditionalization of trials into stereotype-applied vs. -unapplied was conducted individually for each participant on the basis of her or his responses to the items. Accordingly, the design included three primary trial types: nonstereotypical, stereotype-applied, and stereotype-unapplied.

RESULTS

Behavioral Data

On average, participants applied a stereotype on 76.6% of stereotypical trials (e.g., judging that a male target would be very likely to enjoy watching football or would be very unlikely to enjoy shopping for new clothes). Feminine and masculine stereotypes were equally likely to be applied to targets (77.1% and 76.0% of trials, respectively). For nonstereotypical items, participants were equally likely to judge that female and male targets would agree with the statement (52.7% vs. 47.3%, respectively), indicating that participants perceived nonstereotypical items as pertaining equally to females and males. Although nonstereotypical items were judged more slowly ($M = 2170$ msec) than stereotypical items [$M = 2065$ msec; $t(16) = 8.88$, $p = 10^{-7}$], stereotype-applied and stereotype-unapplied judgments were made equally quickly (Ms = 2066 vs. 2064 msec, respectively; $p > .98$).

On the composite IAT posttest measure, participants categorized items more quickly in stereotype-consistent blocks (e.g., female + home/male + career) than stereotype-inconsistent blocks ($M\ diff = 64$ msec) [$t(16) = 2.77$, $p < .01$]. This result replicates a sizeable number of earlier studies demonstrating that, on average, college-aged American perceivers more strongly associate the concepts of careers and science with men and the concepts of home and the humanities with women than vice versa (Nosek et al., 2007; Greenwald, Nosek, & Banaji, 2003; Nosek, Banaji, & Greenwald, 2002). However, we observed substantial variability in the strength of gender associations across participants, such that some individuals demonstrated stronger counter-stereotypical associations (i.e., responding more quickly to blocks in which the concepts of careers and science were paired with women), whereas some individuals demonstrated especially strong stereotypic associations; indeed, the range of composite IAT difference scores ranged from −111 to 304 msec, providing sufficient variability to look for correlations between the behavioral measure and patterns of neural activation.

Confirming its ability to measure the strength of gender-specific associations, the IAT was positively correlated with explicit self-report questions about the importance of gender to participants. The IAT composite score was

most strongly associated with how much more strongly participants judged themselves to be a typical member of their own sex than a member of the other sex [$r(15) = .61, p < .05$]. Composite IAT scores were also significantly correlated with participants' ratings of how important gender was to their identity [$r(15) = .48, p < .05$] and marginally related to how much more similar they perceived themselves to be to the same-sex versus other-sex target [$r(15) = .39, p < .10$].

fMRI Data

To identify the neural correlates of stereotype application, we first examined the contrast of *stereotype-applied > stereotype-unapplied*. This comparison produced an extensive activation in the right frontal cortex (1030 voxels) that included portions of the middle and inferior frontal gyri (Figure 1 displays representative subregions of this right frontal area). This right frontal difference resulted from a significant decrease in right frontal activity during stereotype-unapplied trials relative to both stereotype-applied and nonstereotypical items. A similar, but more circumscribed, pattern of right frontal activation was obtained from the contrast of *stereotype-applied > (stereotype-unapplied + nonstereotypical)*. No brain regions were observed from the contrast of *stereotype-applied + stereotype-unapplied > nonstereotypical*, even at a relaxed statistical threshold of $p < .005$.

Modulation of similar regions was also observed in the left frontal cortex, although the extent of frontal activation in the left hemisphere was considerably less than in the right. In addition, greater activity was observed for *stereotype-applied > stereotype-unapplied* in the left motor cortex, the cingulate, the occipital and parietal cortex, the right superior temporal sulcus, and the bilateral insula (see Table 1). No regions were obtained from the reverse contrast, even at a relaxed statistical threshold of $p < .005$, $k = 5$ voxels.

Correlation between Neural and Behavioral Indices of Stereotyping

We next examined the relation between activation in regions of interest identified from the contrast of *stereotype-applied > stereotype-unapplied* and the strength of participants' stereotypical associations as indexed behaviorally by the IAT. To provide an analysis of potential anatomical differences within the extensive, contiguous set of voxels obtained from this contrast, a set of more circumscribed regions of interest was first identified using an automated search algorithm (R. A. Poldrack, University of California Los Angeles) that defined areas around local maxima separated by at least 8 mm. For each of these subregions, we then calculated the correlation between (i) participants' behavioral index of gender stereotypically, as indexed by the composite IAT difference score and (ii) the difference in BOLD response associated with stereotype-applied versus stereotype-unapplied trials, as indexed by the SPM parameter estimates associated with each trial type (Figure 2).

Significant correlations between behavioral and fMRI indices were observed only in the right frontal cortex. The average Pearson's r across subregions in the right frontal cortex was .39, with a considerable number of loci exceeding the cutoff for a large effect size ($r > .50$, following Cohen, 1988), as well as the critical value for a significant correlation at $p < .05$ ($r = .48$). In contrast, the average correlation in the left frontal cortex was .14, and no left frontal region produced a statistically significant
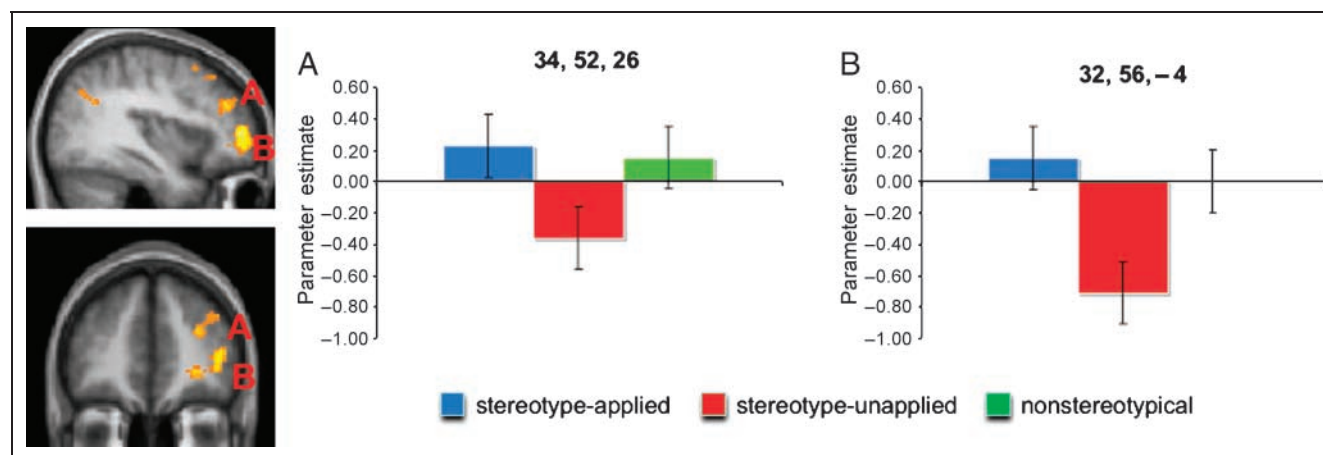

Figure 1. Representative regions of the right frontal cortex obtained from random-effects contrast of *stereotype-applied > stereotype-unapplied*. The extent of activation observed from this contrast included areas of both right middle (A) and right inferior (B) frontal gyri. The pattern of results in these right frontal regions took the form of greater BOLD response to stereotype-applied (leftmost blue bars) and nonstereotypical (rightmost green bars) items than stereotype-unapplied items (middle red bars). Error bars represent the 95% confidence interval for within-subject comparisons.

Table 1. Peak Voxel and Number of Voxels for Regions of Interest Obtained from the Contrast of *Stereotype-applied > Stereotype-unapplied* ($p < .05$, corrected)

Anatomical Label	x	y	z	Voxels	Maximum t
R Frontal cortex	40	50	6	1030	8.30
	32	22	38	161	7.06
	30	10	64	132	5.19
L Motor cortex	−34	−32	48	127	7.35
Cingulate	−10	−26	38	221	6.64
	10	−30	54	38	5.05
	−2	−14	52	46	4.88
	8	12	36	47	4.88
Occipital cortex	−6	−92	−8	57	6.27
Parietal cortex	8	−48	54	99	6.02
	42	−68	38	101	4.67
L Frontal cortex	−28	34	34	167	5.67
R Dorsal STS	58	−28	34	157	5.46
R Insula	34	−8	−4	28	4.68
L Insula	−50	−18	22	29	4.18

t tests reflect the statistical difference between the two conditions, as computed by SPM2.

Coordinates refer to the Montreal Neurological Institute stereotaxic space.

R = right; L = left; STS = superior temporal sulcus.

correlation (see Table 2 for values associated with all frontal subregions). The same lack of a relation between behavioral and fMRI indices of stereotyping was observed in all other regions of interest outside of the right frontal cortex; all *r*s < .25.

DISCUSSION

The current study identified a pattern of neural activation associated with the use of group stereotypes to understand the mind of another person. Participants judged the preferences and opinions of an unfamiliar man and woman, and analyses focused on differences between inferences that were consistent versus inconsistent with widely held gender stereotypes. When participants used a gender stereotype to infer a target's preference (e.g., judging that the female target would enjoy shopping for new clothes or that the male target would not), greater activation was observed in an extensive region of the right frontal cortex, compared to trials when participants judged a target in a manner inconsistent with gender stereotypes. Moreover, the magnitude of this right frontal difference was significantly related to a well-characterized behavioral index of gender stereotyping, the IAT. Specifically, the stronger one's association of men with career/science and of women with home/humanities, the greater the right frontal difference between stereotype-applied and stereotype-unapplied trials. Given that the behavioral measure tapped fairly specific stereotypes about gender roles (home vs. career and science vs. humanities), it is noteworthy that it was significantly related to brain activation in response to a wide variety of stereotype-based inferences. Although a number of additional brain regions distinguished between application and nonapplication of stereotypes (see Table 1), only right frontal cortex modulation was correlated with the behavioral index of stereotyping, suggesting that this region plays an important functional role in the application and suppression of stereotypical inferences about another person's mind.

What might this role be? Right frontal activation has been linked to a surprisingly diverse array of cognitive functions, including semantic retrieval (MacLeod, Buckner,
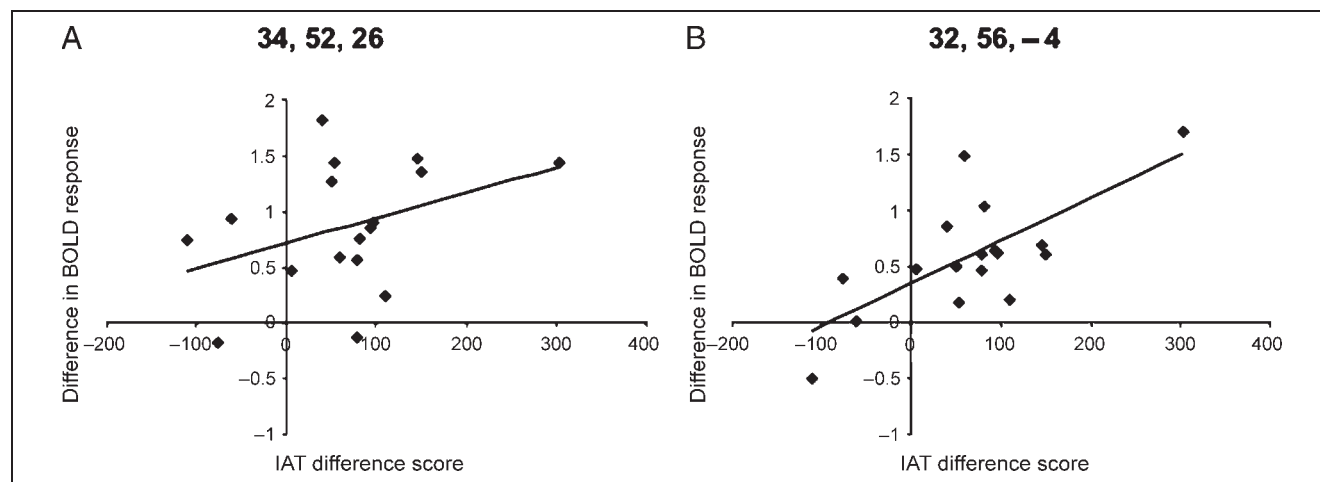

Figure 2. Scatterplots displaying the relation between behavioral and neural measures of gender stereotyping in right middle (A) and right inferior (B) frontal gyri (the same regions as Figure 1). The *x*-axes represent the IAT difference score (larger numbers indicate stronger gender stereotypical associations). The *y*-axes represent the difference between the BOLD response to stereotype-applied and stereotype-unapplied trials. In both regions, a significant correlation between these two measures of stereotyping was observed. Regions correspond to those displayed in Figure 1.

Table 2. Correlation between Stereotypical Gender Associations and the Difference in BOLD Response between *Stereotype-applied > Stereotype-unapplied* in Frontal Subregions

Anatomical Label	x	y	z	Correlation
Right frontal cortex	34	52	26	.71
	32	20	46	.61
	32	42	22	.58
	34	20	54	.58
	32	50	18	.54
	40	14	56	.46
	42	42	30	.46
	22	58	4	.42
	32	56	−4	.38
	40	50	6	.34
	10	60	0	.32
	34	30	34	.29
	30	42	−4	.27
	42	26	48	.27
	32	22	38	.22
	18	56	16	.21
	30	10	64	−.07
Left frontal cortex	−22	62	14	.24
	−38	30	48	.24
	−16	64	8	.23
	−28	38	22	.17
	−32	32	42	.17
	−28	34	34	.10
	−26	16	48	.10
	−30	58	10	.07
	−30	54	26	−.09

Miezin, Petersen, & Raichle, 1998), categorization (Reber, Stark, & Squire, 1998), response inhibition (Aron, Robbins, & Poldrack, 2004), assessment of emotional facial expressions (Nakamura et al., 1999), recognition of one's own face (Keenan, Wheeler, Gallup, & Pascual-Leone, 2000), humor appreciation (Shammi & Stuss, 1999), perception of vocal prosody (Buchanan et al., 2000), and thinking about one's own affective state (Lieberman, 2003). Of these, the current findings may relate most closely to those of MacLeod et al. (1998), who observed right frontal activation during a semantic retrieval task in which participants were asked to consider the category "animals" and monitor for the presentation of the name of a dangerous one. Interestingly, single-cell recordings

in macaques have identified neurons in the frontal cortex that have distinct response profiles to visual stimuli that represent different categories of animals (Freedman, Riesenhuber, Poggio, & Miller, 2001). Consistent with this role of the frontal cortex in categorization, Reber et al. (1998) observed right frontal activation when participants categorized novel visual patterns. Together, these results suggest that the right frontal cortex may play an important role in semantic retrieval of categorical knowledge (e.g., which of the members of the category "animal" are dangerous). Here, we observed similar right frontal modulation when participants made use of a potentially similar aspect of categorical knowledge—that is, stereotypes about social groups—suggesting that this form of social cognition may draw on the same processes as other forms of category-based semantic retrieval.

In contrast, the current results stand apart from the well-regarded conceptual framework developed by Aron et al. (2004), who have suggested that the right frontal cortex plays a critical role in response inhibition, a process important for suppressing or canceling an intended behavioral response. To the extent that participants generally eschew making judgments of others based on their group membership (Vorauer, Hunter, Main, & Roy, 2000; Macrae, Bodenhausen, Milne, & Jetten, 1994; Devine, Monteith, Zuwerink, & Elliot, 1991), one might have expected such response inhibition to accompany trials on which perceivers successfully avoided the use of gender stereotypes. Indeed, a sizeable cognitive literature has demonstrated that stereotypes are often activated automatically upon encountering a social group member, and only prevented from being applied through more explicit control processes (Greenwald & Banaji, 1995; Devine, 1989; Brewer, 1988). Thus, this model suggests that the contrast of *stereotype-applied > stereotype-unapplied* would reveal activity associated with the lack of control. Instead, greater right frontal activation was observed during those trials on which stereotype application was most evident, that is, stereotype-applied trials. Moreover, this pattern of neural activation was strongest for those individuals who demonstrated the most strongly stereotypic associations with gender, that is, among those participants least likely to hold egalitarian beliefs about gender. As such, it seems unlikely that the right frontal activation observed in the current study indexed attempts to inhibit stereotype-consistent responses, but rather seems more likely to represent the application of category knowledge in the service of social judgment (cf. Mason & Macrae, 2004). These data underscore some of the complexities regarding stereotype application and regulation, and suggest the need to revisit existing models of stereotype application through future work.

Interestingly, some earlier studies of mentalizing also observed greater right frontal activity during stories that could be understood only through consideration of the protagonists' mental states (Vogeley et al., 2001; Brunet, Sarfati, Hardy-Bayle, & Decety, 2000; Gallagher et al.,

2000). However, most studies of mentalizing have not observed a modulation of lateral frontal regions (e.g., Saxe & Powell, 2006; Mitchell et al., 2005; Mitchell, Macrae, & Banaji, 2004; Saxe & Kanwisher, 2003; Mitchell et al., 2002), and at least one has reported *less* right frontal activity associated with mentalizing compared to a scrambled-sentence control task (Fletcher et al., 1995). One speculative explanation of these disparate results is that the studies observing right frontal activation during mentalizing have generally used cartoon stimuli that may prompt perceivers to make considerable use of semanticized social knowledge; for example, Brunet et al. (2000) presented participants with a series of highly stylized social situations (e.g., escaping from jail; sending a message in a bottle) that could be understood against the backdrop of other forms of schematic social knowledge. In contrast, most other mentalizing studies present unusual situations or otherwise limit the use of precompiled, semanticized forms of social understanding, which may explain the failure of these studies to observe right frontal activation.

Over the past half-century, psychologists have developed an extensive corpus of work devoted to the role of stereotypes in social cognition and, as such, the relation of the current results to this literature deserves some discussion. First, the primary focus of the current study has been the way that perceivers may deploy stereotypes for a specific purpose: making inferences about another person's mind. In contrast, many of the stereotypes that exist about social groups cannot as readily aid perceivers' attempts at mentalizing, such as those that pertain to physical abilities (e.g., men as strong, African-Americans as athletic) or occupational roles (e.g., women as nurses, Jews as bankers). Additional research will be needed to determine whether the current results are specific to those stereotypes that provide insight into the putative mental characteristics of other people or extend to those that instead relate to nonmental aspects of intergroup beliefs, a distinction rarely drawn in the existing cognitive literature on stereotyping.

Second, the current results specifically contrast those situations in which perceivers apply a stereotype in the process of understanding another's mind to those in which they mentalize on some other basis. Although a substantial amount of research has demonstrated that encountering a member of a minority group (or verbal labels denoting such a group) often leads to the *activation* of associated stereotypical beliefs about the group, the actual application of a stereotype to a specific individual may rely on different cognitive processes than mere activation of stereotypic content (Gilbert & Hixon, 1991; Devine, 1989). That is, although a perceiver may become aware of a host of stereotypical associations upon encountering a member of an outgroup, the use of that content for drawing inferences about the person (i.e., its application) may or may not ensue. The current results examine the processes associated with the phase of social judgment at which perceivers are either applying or avoiding a stereotype, rather than the activation phase that presumably precedes such application. Likewise, the current study differs from those that ask perceivers to report explicitly on the content of stereotypes; for example, by considering whether a given behavior is generally held by others to be more likely of men or women (Quadflieg et al., in press).

Third, the current study should be distinguished from earlier research on the neural basis of prejudice, which has primarily focused on *attitudes* toward—that is, the positive or negative evaluations of—members of an outgroup (Knutson, Mah, Manly, & Grafman, 2007; Amodio & Devine, 2006; Cunningham et al., 2004; Cunningham, Johnson, Gatenby, Gore, & Banaji, 2003; Phelps, Cannistraci, & Cunningham, 2003; Hart et al., 2000; Phelps et al., 2000) rather than stereotypes. Although stereotypes may connote a particular evaluation, one's stereotypes about a group need not converge on a single attitude toward that group; women are stereotypically considered weak but caring, whereas men are considered strong but emotionally stunted. In the current study, we were careful to measure stereotypes about men and women that were equally positive, both on the behavioral IAT measure (science vs. humanities; home vs. career) and on the judgment task (e.g., like baseball vs. like the ballet).

Although much of this earlier work on prejudice has identified the amygdala as the brain region most reliably modulated by the consideration of outgroup versus ingroup members, more recent work has highlighted the contributions of additional brain regions during social evaluation. For example, Knutson et al. (2007) examined brain activation while participants completed blocks of the IAT that were either consistent or inconsistent with prevailing race attitudes toward Black and White Americans. Interestingly, these authors observed a correlation between the strength of participants' race attitudes (as measured by the IAT) and modulation of the BOLD response in a very similar right frontal locus as the one reported in the current study. Curiously, this correlation was not observed for an IAT designed to measure gender stereotyping; indeed, these authors report observing no differences that were specific to stereotyping (instead, all reported differences appear to have been driven by the attitude measure or general task demands of the IAT on executive function). These observations of both overlap and dissociation between the current study and that of Knutson et al. provide ample impetus for future research designed to examine whether stereotyping shares some cognitive processes with purely evaluative aspects of social judgment.

Lastly, established models of stereotyping have suggested that perceivers can deploy two different sets of person perception processes. One of these makes use of information about a target's membership in various social groups, especially those that are immediately obvious to a perceiver such as a person's age, race, and sex

(Brewer, 1988). Stereotyping results directly from this kind of group-based approach to other people (Fiske & Neuberg, 1990; Allport, 1954). In contrast, perceivers may also make use of individuating information about others, which is typically thought to involve a tradeoff between ease of processing and the accuracy of one's interpersonal judgments; stereotypes may be "fast and frugal" heuristics for making inferences about others, but they are generally less specific and contentful than judgments that follow a moment of individual consideration. Recently, these insights into the nature of person perception have begun to be linked to the activity of specific brain regions involved in social cognition. Whereas the type of mentalizing associated with the medial prefrontal cortex may involve more individual consideration of another person's mind, the heuristic application of stereotypes to understand others may rely on the right lateral frontal cortex. Consistent with this formulation, we recently reported modulation of a region of the ventromedial prefrontal cortex when perceivers mentalized about an unfamiliar target who had previously been considered from a first-person perspective (Ames, Jenkins, Banaji, & Mitchell, in press), a manipulation known to reduce stereotype-consistent judgments of others (Ames, 2004a, 2004b; Galinsky & Moskowitz, 2000).

As such, the current results contribute to an emerging consensus that, rather than relying on a single module, the human ability to understand other minds draws on a suite of distinct cognitive processes, each of which relies on different kinds of information to solve the overall challenges posed by mentalizing. One useful mechanism for making inferences about others may be to use knowledge of one's own thoughts, feelings, and preferences as a guide to those of others (Gallese, 2007; Davies & Stone, 1995). However, although this strategy of simulation (or projection) may be appropriate for understanding the mental states of someone assumed to share one's own worldview (Mitchell et al., 2005, 2006), life in the 21st century includes frequent encounters with individuals who may think in ways very different from our own. How, then, do perceivers mentalize when they assume that a target fails to share their own predilections and proclivities and therefore cannot appropriately be simulated? Here, we suggest that one possible alternative to individuating other minds through simulation is to base judgments on "precompiled" semantic beliefs about the social categories to which a target belongs, that is, on a stereotype. Although such stereotypical inferences may prove inaccurate guides to an individual's true thoughts and feelings (Jussim, 1991, 1993), perceivers, nevertheless, frequently deploy categorical social knowledge during their attempts to make sense of others. The undesirable results of stereotyping may result from the fact that, rather than drawing on cognitive processes specialized for thinking about the minds of others (Harris & Fiske, 2006; Mitchell, 2006;

Mitchell et al., 2002), stereotype-based social judgments rely on brain regions—such as the right lateral frontal cortex—that subserve more general-purpose cognitive processes, such as those involved in semantic memory and categorization. By beginning to illuminate the neural basis of these less individuating forms of mentalizing, the current results underscore the diverse cognitive mechanisms of which the human mind makes use in its attempt to understand the complexity of the other minds around it.

REFERENCES

Allport, G. (1954). *The nature of prejudice.* Cambridge: Addison-Wesley.

Ames, D. L., Jenkins, A. C., Banaji, M. R., & Mitchell, J. P. (in press). Taking another's perspective increases self-referential neural processing. *Psychological Science.*

Ames, D. R. (2004a). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87,* 340–353.

Ames, D. R. (2004b). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology, 87,* 573–585.

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91,* 652–661.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–177.

Brewer, M. B. (1988). A dual model of impression formation. In T. K. Srull & R. S. J. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Brunet, E., Sarfati, Y., Hardy-Bayle, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage, 11,* 157–166.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Brain Research, Cognitive Brain Research, 9,* 227–238.

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 NY Academy of Sciences, 1124,* 1–38.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12,* 1–47.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2,* 913–919.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Cunningham, W. A., Johnson, M. K., Gatenby, J., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85,* 639–649.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Chris Gatenby, J., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science, 15,* 806–813.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8,* 109–114.

Davies, M., & Stone, T. (Eds.). (1995). *Mental simulation: Evaluations and applications*. Oxford, UK: Blackwell Publishers.

Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Devine, P. G., Monteith, M., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60,* 817–830.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87,* 327–339.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology, 23,* 1–74.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of ''theory of mind'' in story comprehension. *Cognition, 57,* 109–128.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291,* 312–316.

Gabrieli, J. D. E., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 906–913.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78,* 708–724.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ''theory of mind''. *Trends in Cognitive Sciences, 7,* 77–83.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of ''theory of mind'' in verbal and nonverbal tasks. *Neuropsychologia, 38,* 11–21.

Gallese, V. (2007). Before and below ''theory of mind'': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362,* 659–669.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60,* 509–517.

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport, 6,* 1741–1746.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Hamilton, D. (1981). *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, NJ: Erlbaum.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science, 17,* 847–853.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *NeuroReport, 11,* 2351–2355.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, U.S.A., 96,* 9379–9384.

Jussim, L. (1991). Social perception and social reality: A reflection–construction model. *Psychological Review, 98,* 54–73.

Jussim, L. (1993). Accuracy in interpersonal expectations: A reflection–construction analysis of current and classic research. *Journal of Personality, 61,* 637–668.

Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal Psychology and Social Psychology, 28,* 280–290.

Keenan, J. P., Wheeler, M. A., Gallup, G. G., Jr., & Pascual-Leone, A. (2000). Self-recognition and the right prefrontal cortex. *Trends in Cognitive Sciences, 4,* 338–344.

Knutson, K. M., Mah, L., Manly, C. F., & Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Human Brain Mapping, 28,* 915–930.

Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Explicit and implicit processes* (pp. 44–67). New York: Cambridge University Press.

MacLeod, A. K., Buckner, R. L., Miezin, F. M., Petersen, S. E., & Raichle, M. E. (1998). Right anterior prefrontal cortex activation during semantic monitoring and working memory. *Neuroimage, 7,* 41–48.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66,* 37–47.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology, 67,* 808–817.

Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron, 31,* 653–660.

Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. *Journal of Cognitive Neuroscience, 16,* 1785–1795.

Mitchell, J. P. (2005). The false dichotomy between simulation and theory–theory: The argument's error. *Trends in Cognitive Sciences, 9,* 363–364; author reply 364.

Mitchell, J. P. (2006). Mentalizing and Marr: An information processing approach to the study of social cognition. *Brain Research, 1079,* 66–75.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17,* 1306–1315.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, U.S.A., 99,* 15238–15243.

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,* 4912–4917.

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

Nakamura, K., Kawashima, R., Ito, K., Sugiura, M., Kato, T., Nakamura, A., et al. (1999). Activation of the right inferior frontal cortex during assessment of facial emotion. *Journal of Neurophysiology, 82,* 1610–1614.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125,* 737–759.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop: Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition and Emotion, 15,* 853–864.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, Å. H. (2000). Emotional state and the detection of change in facial expression of emotion. *European Journal of Social Psychology, 30,* 211–222.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6,* 101–115.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18,* 36–88.

Phelps, E. A., Cannistraci, C. J., & Cunningham, W. A. (2003). Intact performance on an indirect measure of face bias following amygdala damage. *Neuropsychologia, 41,* 203–208.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729–738.

Quadflieg, S., Turk, D. J., Waiter, G. D., Mitchell, J. P., Jenkins, A. C., & Macrae, C. N. (in press). Stereotyping modulates reward circuitry. *Journal of Cognitive Neurosciences.*

Reber, P. J., Stark, C. E., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 747–750.

Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences, 9,* 174–179.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage, 19,* 1835–1842.

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science, 17,* 692–699.

Shammi, P., & Stuss, D. T. (1999). Humour appreciation: A role of the right frontal lobe. *Brain, 122,* 657–666.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., et al. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage, 14,* 170–181.

Vorauer, J. D., Hunter, A. J., Main, K. J., & Roy, S. A. (2000). Meta-stereotype activation: Evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology, 78,* 690–707.

Wagner, A. D., Schacter, D. L., Rotte, M., Koustaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experience as predicted by brain activity. *Science, 281,* 1188–1191.