How Do You See Me? The Neural Basis of Motivated Meta-perception

Taru Flagan[1], Jeanette A. Mumford[2], and Jennifer S. Beer[1]

Abstract

■ We cannot see the minds of others, yet people often spontaneously interpret how they are viewed by other people (i.e., meta-perceptions) and often in a self-flattering manner. Very little is known about the neural associations of meta-perceptions, but a likely candidate is the ventromedial pFC (VMPFC). VMPFC has been associated with both self- and other-perception as well as motivated self-perception. Does this function extend to meta-perceptions? The current study examined neural activity while participants made meta-perceptive interpretations in various social scenarios. A drift-diffusion model was used to test whether the VMPFC is associated with two processes involved in interpreting meta-perceptions in a self-flattering manner: the extent to which the interpretation process involves the preferential accumulation of evidence in favor of a self-flattering interpretation versus the extent to which the interpretation process begins with an expectation that favors a self-flattering outcome. Increased VMPFC activity was associated with the extent to which people preferentially accumulate information when interpreting meta-perceptions under ambiguous conditions and marginally associated with self-flattering meta-perceptions. Together, the present findings illuminate the neural underpinnings of a social cognitive process that has received little attention to date: how we make meaning of others' minds when we think those minds are pointed at us. ■

INTRODUCTION

Recent research has extended the role of ventromedial pFC (VMPFC) in social cognition to processing that is shaped by certain socioemotional motivations (e.g., the desire to see oneself in a flattering light: Delgado et al., 2016; Flagan & Beer, 2013; Hughes & Beer, 2013; Beer, 2007), yet its role in meta-perceptions has not been fully characterized. Very little is understood about the neural underpinnings of meta-perceptions, that is, how we interpret others' thoughts about us. A few studies have examined the neural associations of meta-perceptions, and most have found an association with VMPFC activity (e.g., Veroude, Jolles, Croiset, & Krabbendam, 2014; Pfeifer et al., 2009; Ochsner et al., 2005, but see also D'Argembeau et al., 2007). Yet, the psychological significance of the neural associations with meta-perceptions is not known, and no attention has been paid to the motivated nature of interpreting other people's thoughts about the self. For example, people are likely to strive toward self-flattering interpretations of other people's thoughts particularly when the only available information is ambiguous (Preuss & Alicke, 2009; Sedikides & Gregg, 2008; Dunning, Meyerowitz, & Holzberg, 1989; Markus, Smith, & Moreland, 1985). The current study utilized a model-based approach to examine the neural underpinnings of (a) the expectations and preferential information processing that can contribute to interpreting others' minds when available information is normatively ambiguous and (b) self-flattering interpretations.

Previous research suggests that people often set out to see themselves in a positive light and expect that others will also think well of them (e.g., Preuss & Alicke, 2009; Markus et al., 1985). How can the influence of this motivation be measured? In situations where meta-perceptions are likely to be positive (e.g., interactions with a close other or clear cues of the person's positive thoughts about you), there is a match between the goal of the perceiver and the normative meta-perception making it difficult to distinguish between the two. In other words, a self-flattering interpretation in a positive scenario might arise because that is a reasonable interpretation given the available information, or it might arise because distorted processing is used to ensure a self-flattering interpretation.

A better measurement of self-flattering meta-perceptions is the comparison of meta-perceptions of scenarios in which meta-perceptive targets' behavior toward the self is relatively ambiguous versus negative. People tend to engage in similar evidence accumulation processes when interpreting others' minds in relatively ambiguous and negative scenarios; they strike a compromise between their desired interpretation and plausible interpretations when sifting through evidence (Ditto & Lopez, 1992; Dunning et al., 1989). Although the desired interpretation remains constant across conditions,

[1]University of Texas at Austin, [2]University of Wisconsin, Madison

Journal of Cognitive Neuroscience 29:11, pp. 1908–1917
doi:10.1162/jocn_a_01169

meta-perception in ambiguous situations lends itself to a wider range of plausible interpretations (some of which will be self-flattering) than negative situations. On average, a thorough analysis of all plausible interpretations in a relatively ambiguous scenario would require at least as much consideration than considering the smaller range of plausible interpretations (which tend to be treated with skepticism: Ditto & Lopez, 1992) in the case of negative scenarios. However, the exploitation of plausible self-flattering interpretations for meta-perceptions in ambiguous scenarios can actually result in shallower evidence accumulation in comparison with interpreting others' minds in scenarios that are more straightforwardly negative (e.g., Beer & Hughes, 2010; Dunning et al., 1989). Therefore, the extent to which people interpret someone's mind in a self-flattering matter can be reflected by individual differences in the extent to which they require shallower evidence accumulation to reach self-flattering meta-perceptive interpretations in relatively ambiguous scenarios than in relatively negative scenarios.

Although it has not been directly studied, the VMPFC is the most likely neural candidate to mediate meta-perceptions that are of a self-flattering nature. Whereas both medial pFC and the TPJ have been associated with general mentalizing (for a review, see Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), it is the VMPFC that has been implicated in self-evaluations and meta-perceptions (e.g., Veroude et al., 2014; Jenkins & Mitchell, 2011; Pfeifer et al., 2009; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006; Ochsner et al., 2005; Kelley et al., 2002) including self-evaluations of a self-protective nature (e.g., Chavez, Heatherton, & Wagner, 2016; Hughes & Beer, 2012, 2013). Furthermore, self-flattering meta-perceptions are most likely to be evident in ambiguous social scenarios (Preuss & Alicke, 2009; Dunning et al., 1989; Markus et al., 1985). VMPFC has been associated with mentalizing about others' minds in conditions of ambiguity (i.e., uncertainty that is one kind of ambiguity: Jenkins & Mitchell, 2010) and the extent to which ambiguous facial expressions of other people are seen as positive (Kim, Somerville, Johnstone, Alexander, & Whalen, 2003). However, no previous research has tested the underlying psychological role of VMPFC in meta-perceptions or its association with self-flattering meta-perceptions.

Although VMPFC is the strongest candidate, it is also important to test for the possibility that the amygdala and ACC are associated with motivated meta-perceptions. Although they have not tended to show associations with mind perception, both amygdala and ACC have been associated with self-protective evaluations or flattering evaluations of others. For example, amygdala activity is shaped by the goal of person evaluation such that it is associated with positive evaluations of people whom you are motivated to see in a positive light (Schiller, Freeman, Mitchell, Uleman, & Phelps, 2009; Cunningham, Van Bavel, & Johnsen, 2008). The dorsal ACC has been associated with monitoring for instances in which one might look foolish (Bengtsson, Dolan, & Passingham, 2011). However, as in the case of the VMPFC, it remains unknown whether the amygdala or ACC is associated with interpreting the minds of others in a self-flattering manner.

The current study utilized drift-diffusion modeling (DDM; Ratcliff, 1978) and fMRI to examine functional neural networks associated with the evidence accumulation processes underlying meta-perception and self-flattering interpretations of others' minds. It is difficult to distinguish expectations from evidence accumulation using self-reported choice or RTs; both might lead to a particular choice or a faster RT. DDM independently estimates preferential evidence accumulation (i.e., drift rate) from expectations (i.e., starting points). Furthermore, DDM includes parameter estimates that address extraneous processes more broadly encompassed by choice and RT measures. In other words, DDM makes it possible to test whether meta-perceptions are accomplished through a selective evidence accumulation process that favors a particular interpretation versus an a priori expectation about the self-flattering nature of someone's thoughts without regard to the specific content of a scenario.

Participants were asked to place themselves in positive, negative, and ambiguous social scenarios and then chose whether self-flattering or non-self-flattering interpretations best characterized how other people were likely to be thinking about them. Although most people are motivated to see themselves in a positive light, they are not delusional and are expected to calibrate their expectations to the most self-protective degree possible while remaining anchored in reality (e.g., Taylor & Brown, 1988). In a forced choice task (i.e., self-flattering vs. non-self-flattering options), a lack of bias would predict a starting point right around the midpoint. Therefore, participants were expected to move their starting points in a flattering direction in the positive condition but conform to the base rate of the task when expectations of flattery would be too far removed from the constraints of the situation (i.e., expectations for the negative and ambiguous should be around the midpoint). People tend to take more time to consider evidence that skews negatively about themselves (e.g., Ditto & Lopez, 1992) and need more time on average to sort through the conflicting information contained in ambiguous scenarios (e.g., Dunning et al., 1989). Therefore, although the underlying reasons may be different, previous research predicts that both the negative and ambiguous conditions should be subject to slower evidence accumulation (than the positive condition) and not necessarily different from one another on average. Neuroimaging analyses sought to identify which, if any, of the hypothesized neural regions are associated with the evidence accumulation processes that underlie meta-perceptions including those of a self-flattering nature.

METHODS

Participants

Sixty-two participants underwent fMRI while reporting on their meta-perceptive interpretations of social scenarios. One participant was excluded from data analyses for failure to respond on 58% of the trials, six additional participants were excluded from analyses because of excessive head movement (>3 mm), and one participant was excluded for a structural abnormality. Data analysis focused on the remaining 54 participants (27 women; mean = 21.04 years, *SD* = 2.79 years). All participants provided informed consent, and the study was approved by the institutional review board of the University of Texas at Austin. Participants were compensated $15/hr or research credit for their participation. All participants were right-handed, fluent English speakers, and free from medications and neurological conditions that might influence the measurement of cerebral blood flow and had normal or corrected-to-normal vision.

Procedure

The experimental procedure was designed to use DDM to test the role of expectations and evidence accumulation in meta-perceptions of social scenarios, which requires a forced choice format (e.g., Voss & Voss, 2007; Voss, Rothermund, & Voss, 2004; Ratcliff, 1978). Participants were presented with social scenarios and instructed to imagine themselves in that scenario as the object of someone else's attention (see Figure 1). Participants were then asked to choose a meta-perception that best characterized how they would interpret others' actions in the scenarios to reflect on themselves (i.e., reflecting on the participants). The social scenarios included scenarios where meta-perceptions were normatively interpreted as negative, positive, or relatively ambiguous toward the participant (Negative, Positive, and Ambiguous conditions, respectively). Analyses focused on the Ambiguous and Negative conditions; the Positive condition ensured a full range of scenarios for participants. The inclusion of only the Ambiguous and Negative

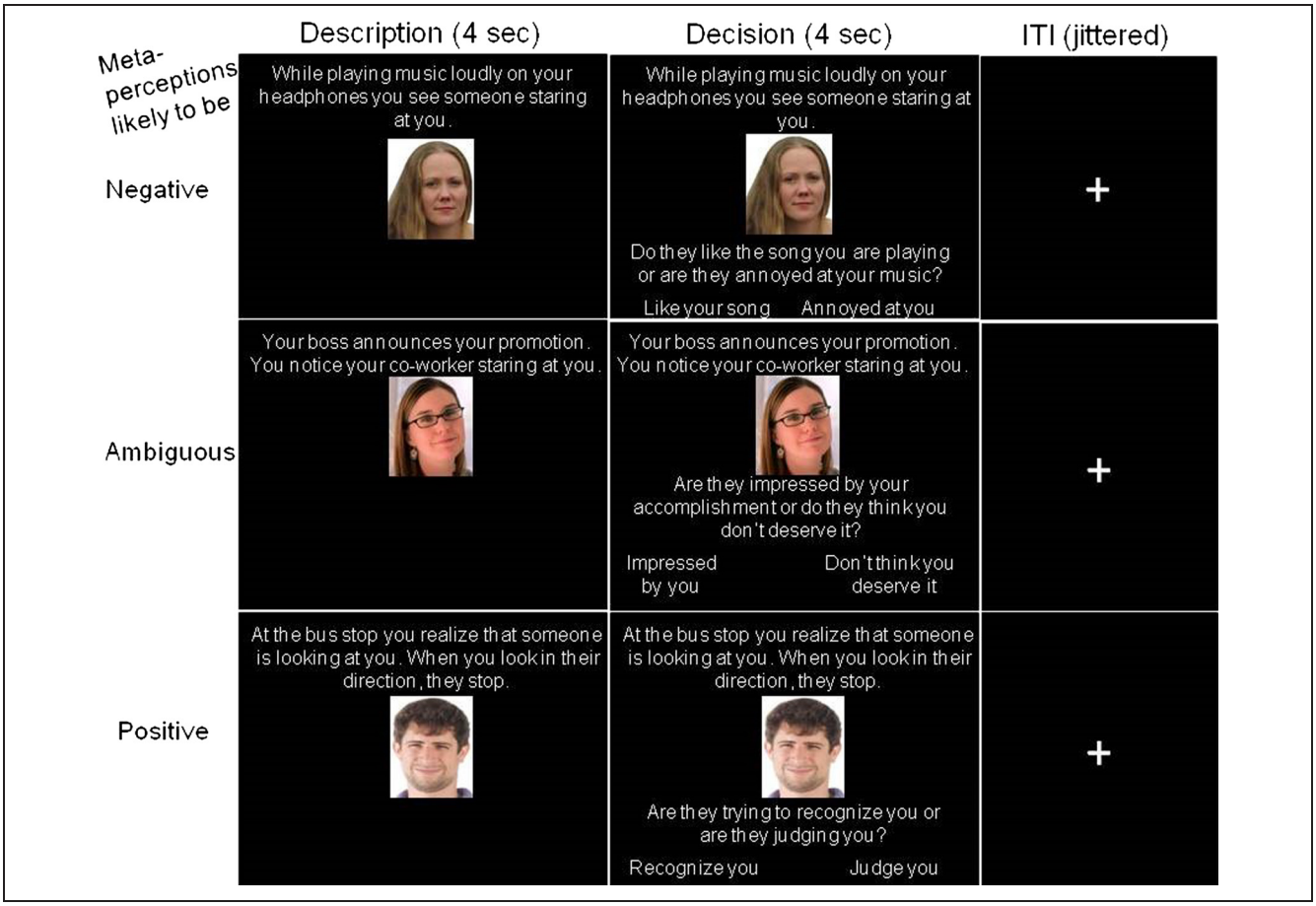

Figure 1. Meta-perception task. Each trial presented the scenario (4000 msec) followed by a question and two response options about how others in the situation would be thinking about the participant (4000 msec). The self-flattering and non-self-flattering response options were counterbalanced such that the flattering option was the first option presented on 50% of the trials within a condition and the second option on 50% of the trials within a condition. The general content of the scenario, length of statement and questions, and sociality of the pictures and statements of the ambiguous scenarios were matched with the negative and positive scenarios. ITI = intertrial interval.

conditions would make it difficult to know whether individual differences between the two conditions arose simply because the Ambiguous condition seemed more flattering in contrast. For each scenario, participants were asked to choose whether a flattering or unflattering interpretation was most descriptive of how they would perceive others to view them in that scenario. Flattering options involved choices that were either self-enhancing (i.e., exaggerated positive qualities of the self; 25% in each condition) or self-protective (i.e., did not inflate positive qualities but instead gave a relatively neutral option that was at least not threatening to the self; 75% in each condition; see Alicke & Sedikides, 2009). The interpretation options were worded such that it was clear that participants were rating how they perceived others to perceive them (i.e., meta-perceptions; e.g., "While playing music loudly on your headphones you see someone staring at you. Do they like the song you are playing or are they annoyed at the loudness of your music?").

The social scenario stimuli used in the fMRI study were developed by the authors for the purpose of this experiment. Initial pilot testing involved 267 stimuli and numerous samples of judges. As in previous research on motivated interpretation, stimuli were categorized as ambiguous if judges, on average, were only as likely to select the self-flattering answer option as the non-self-flattering answer option (i.e., ambiguity arises from lack of social consensus rather than uncertainty: Neta, Kelley, & Whalen, 2013; Neta & Whalen, 2010; Beard & Amir, 2009; Neta, Norris, & Whalen, 2009). These pilot tests were used to reduce the stimuli to a planned set of 112 items. The final pilot test was restricted to ratings of only the planned 112 items, which were judged by 114 judges (79 women; mean age = 19.06 years, SD = 2.25 years) drawn from the same participant pool used for the fMRI experiment. None of the judges took part in the fMRI experiment. All stimuli in both the pilot testing and the fMRI experiment included descriptions of social scenarios paired with photos and were followed by a forced choice of how the scenario reflected on the self.

The general content of the scenario, length of statement and questions, and sociality (group vs. one other person) of the social scenario pictures and statements of the ambiguous scenarios were matched with the negative and positive scenarios. Pictures were included to increase participant engagement and provide a similar visual starting point for considering the scenario. In each condition, about a quarter of the photographs did not depict a person (e.g., showed a text or email message), and about 10% depicted groups of men and women or did not include any facial expression (e.g., depicted a person from the back). For the photographs that did depict a single gender, most were of women (72% for Ambiguous, 78% for Negative, and 76% Positive). For the photographs that did include at least one facial expression, they tended to be neutral facial expressions (39.4% Ambiguous, 43% Negative, and 11% Positive) or smiles (29% Ambiguous, 33% Negative, and 68% Positive). There was no significant empirical evidence that planned comparisons of participants' responses across conditions could be accounted for by gender or facial expression of photographs (i.e., interaction terms for Condition × Gender and Condition × Face for response and RT ranged from 0.01 to 1.22, ps ranged from .37 to .93).

When possible, a similar scenario was used across all three conditions. For example, the meta-perceptive interpretation of someone's stare was used in all three conditions: Negative condition: "While playing music loudly on your headphones you see someone staring at you. Do they like the song you are playing or are they annoyed at the loudness of your music?"; Ambiguous condition: "Your boss announces your promotion at work. You notice your co-worker staring at you. Are they impressed by your accomplishment or do they think you don't deserve it?"; Positive condition: "At the bus stop you realize that someone is looking at you. When you look in their direction, they stop. Are they trying to recognize you or are they judging you?" Scenarios that did not lend themselves to all three conditions were always matched between the Ambiguous condition and either the Negative or Positive condition. For example, the meta-perceptive interpretation of texting someone and hoping for a response was used in both the Negative and Ambiguous conditions (i.e., Negative condition: "You text your crush asking if they want to go the game this evening. You see that they read the message but did not reply. Do they not want to go to the game with you or have they not had time to reply?"; Ambiguous condition: "You invite a new acquaintance out for coffee but receive no reply. Is your new acquaintance having second thoughts about spending time with you or busy?"). The meta-perceptive interpretation of receiving a thumbs up was used in both the Positive and Ambiguous conditions (i.e., Positive condition: "On the first day of an assigned group project you make a suggestion about the project. One of your group members gives you a 'thumbs up.' Are they being sarcastic or supportive?"; Ambiguous condition: "During lecture, you ask your professor a question. While your professor is thinking about how to respond, a classmate gives you a 'thumbs up.'").

In the pilot test of the stimuli set used in the present research, repeated-measures ANOVAs found main effects of Condition for both response times ($F(2, 226) = 59.62$, $p < .001$) and ratings ($F(2, 226) = 808.20$, $p < .001$). Judges took the longest to rate the Ambiguous scenarios (mean = 2172.95 msec, SD = 359.31 msec) compared with the Negative scenarios (mean = 2053.78 msec, SD = 366.52 msec; $t(113) = 7.97$, $p < .05$) and the Positive scenarios (mean = 2012.39 msec, SD = 382.14 msec; $t(113) = 11.39$, $p < .05$). The ambiguous scenarios had an average rating of 0.54 (SD = 0.13; negative responses were coded as a 0; positive responses were coded as a 1), the negative scenarios had an average rating of 0.28 (SD = 0.13), and the positive scenarios had an average rating of 0.75 (SD = 0.14).

In the fMRI experiment, participants performed the meta-perception task (see Figure 1). For each trial, both of the self-flattering and non-self-flattering response options were always presented simultaneously, and responses were counterbalanced such that the flattering option was the first option presented on 50% of the trials within a condition and the second option on 50% of the trials within a condition. Participants indicated their meta-perception choice by pressing a button that corresponded to one of the two answer options. After a response was made, a fixation screen appeared until the end of the trial. The experiment consisted of 56 ambiguous scenario trials (Ambiguous condition), 28 negative scenario trials (Negative condition), and 28 positive scenario trials (Positive condition). Trials were separated by fixation (crosshair) screens so that neural activation associated with the social scenario descriptions could be independently modeled. Participants were instructed to clear their minds during fixation screens. Fixations were presented at a jittered time drawn randomly from a truncated exponential distribution (intertrial interval: mean = 3 sec, max = 8 sec; e.g., Mumford, Turner, Ashby, & Poldrack, 2012; Dale, 1999). Trials from Ambiguous, Negative, and Positive conditions were counterbalanced and pseudorandomized across four runs of 4 min 44 sec each.

Fitting the DDM to the Data

Diffusion model data analysis was conducted with *fastdm* (Voss & Voss, 2007). DDM is a variant of continuous sampling models proposed for two-alternative forced-choice decisions (Ratcliff, 1978). Although originally applied to decisions in which choices could be classified as correct or incorrect, recent research has shown the applicability of DDMs to decisions about ambiguous stimuli in which each choice is equally applicable (Germar, Albrecht, Voss, & Mojzisch, 2016; Voss & Schwieren, 2015; Voss, Rothermund, & Brandtstädter, 2008) and to decisions that reflect subjective preferences that cannot be reduced to correct or incorrect choices (Krajbich, Lu, Camerer, & Rangel, 2012; Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010). DDM is fit to the response and RT data.

In DDM, the model assumes that, when deciding between two options, information from a stimulus is sampled over time, beginning from an initial value called the starting point (z), until a decision boundary is reached (a or 0) and a decision response is initiated. The relation of the starting point to the upper threshold (z/a) reflects the expectations that precede the decision process. For example, if z is closer to the upper threshold a, this suggests a prior expectation of outcome a, whereas if z is closer to the lower threshold 0, this suggests a prior expectation of outcome 0. In addition, the rate at which information is accumulated toward a decision is measured by the drift rate (v), which reflects the strength of decision evidence. Faster drift rates indicate facilitated, that is, shallower processing of information. The difference in nondecisional time parameter (d) captures the mean difference in nondecisional time for responses corresponding to the upper threshold and the lower threshold. Finally, DDM estimates four other parameters: a parameter for nondecision processes (*t0*) and three parameters that index trial variability (variability in starting point, *sz*; variability in drift rate, *sv*; and variability in nondecisional components *st0*; Ratcliff, 1978). The current study focused on starting point and drift rate. Starting point and drift rate have dissociable effects in terms of the RT distribution's shift and skew as well as choice probabilities (White & Poldrack, 2014). For example, simulations and empirical investigations have shown that starting point bias leads to a shift in the leading edge of the RT distribution (e.g., the fastest responses as assessed by the 0.2 quantile), whereas drift rate bias does not. Instead, drift rate biases are more likely to affect the skew of the RT distribution (i.e., both fast and slow responses; White & Poldrack, 2014).

To investigate the neural regions supporting the starting points and rates of evidence accumulation that contribute to interpreting meta-perceptions, a DDM was fit to the response and RT data, and the eight parameters mentioned above were estimated. The upper threshold (a) corresponds to the self-flattering interpretation of a social situation, and the lower threshold (0) corresponds to the non-self-flattering interpretation of a social situation. The starting point (z), drift rate (v), and difference in nondecisional time (d) were estimated for the three conditions (Ambiguous, Negative, and Positive), holding all other model parameters (a, *t0*, *sz*, *sv*, *st0*) constant (Voss & Voss, 2007; Voss et al., 2004). The starting point, relative to the upper threshold (a), captured preexisting expectations of social situations by estimating how much the starting point favored one decision (e.g., self-flattering interpretation) over the other (e.g., non-self-flattering interpretation) for ambiguous, negative, and positive social situations. The drift rate captured sensitivity to information in social situations by estimating the rate at which information is accumulated toward a decision threshold in each condition. The difference in nondecisional time (d) captures the mean difference in nondecisional time for responses corresponding to the upper threshold (i.e., self-flattering interpretation) and the lower threshold (i.e., non-self-flattering interpretation). Model fit was assessed using the Kolmogorov–Smirnov (KS) statistic, which is robust against outliers and uses the entire empirical distributions of RTs (Voss & Voss, 2007; Voss et al., 2004). For each model, we computed the mean probability value of the KS statistic when comparing the empirical distribution with the predicted distribution. Small probability values (e.g., $p < .05$) for the KS statistic indicate significant deviations between the empirical and predicted distributions.

Behavioral Analysis

A repeated-measures ANOVA tested whether condition had a significant effect on the DDM parameter estimates of drift rates (absolute value), starting points, and differences in nondecisional time. Higher values of drift rate indicate more greatly facilitated evidence accumulation. Starting points closer to 1 indicate expectations that meta-perceptions will be self-flattering, and starting points closer to 0 indicate expectations that meta-perceptions will be non-self-flattering. Positive values of differences in nondecisional time indicate faster RTs for responses corresponding to the upper threshold than responses corresponding to the lower threshold. Paired t tests were conducted to test for significant differences in RT and responses across the three conditions.

fMRI Data Acquisition and Preprocessing

Imaging data were acquired on a 3-T Skyra MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. Functional data were collected using a T2*-weighted EPI sequence (repetition time = 2000 msec, echo time = 30 msec, flip angle = 63°, field of view = 230, voxel size = 2.4 × 2.4 × 2.4) and time-locked to initial trial onset. Fifty-six axial slices were positioned 30° off the AC–PC line to reduce frontal signal dropout (Deichmann, Gottfried, Hutton, & Turner, 2003). Slices were acquired using the multiband sequence (Moeller et al., 2010; acceleration factor = 2, parallel imaging factor iPAT = 2) in an interleaved fashion. Higher-order shimming was used to reduce susceptibility artifacts. A high-resolution full-brain image using a magnetization prepared rapid gradient echo pulse sequence (repetition time = 1900 msec, inversion time = 900 msec, echo time = 2.43 msec, flip angle = 9°, field of view = 256) was acquired for image registration.

Neuroimaging data were preprocessed and analyzed using the FSL software toolbox (Oxford Center for Functional MRI of the Brain; Smith et al., 2004). Raw imaging data were converted from DICOM format to NIFTI format. Functional images were motion corrected using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), and non-brain structures were stripped from functional and structural volumes using the Brain Extraction Tool (Smith, 2002). Low-frequency noise was removed using a high-pass filter of a Gaussian-weighted least-squares straight fit line with a cutoff of 100 sec. Data were resampled to 2-mm cubic resolution, and spatial smoothing was performed using a Gaussian kernel with an FWHM of 5 mm. Data were first registered to the high-resolution T1-weighted structural image using Boundary-Based Registration, which was then registered to the standard brain (MNI152 2-mm template) using 12 DOF affine registration.

fMRI Analysis

For each participant and each run, a general linear model (GLM) was estimated in FSL's FEAT (fMRI Expert Analysis Tool Version 5.98) first level analysis package with (a) three regressors (ambiguous, negative, and positive) convolved with a canonical double-gamma hemodynamic response function and temporal derivative and (b) seven regressors of noninterest to account for missed trials and the six directions of head movement. A GLM analysis created a contrast image of interest for each participant (e.g., Ambiguous > Negative). A second level analysis (fixed effects) was conducted to average the contrasts across the four runs for each participant. The averaged scans were entered into a group level random effects analysis using Oxford Center for Functional MRI of the Brain's Local Analysis of Mixed Effects (Smith et al., 2004).

Group level analyses were limited to hypothesized neural regions. Within the relevant neuroanatomical ROIs in the automated anatomical labeling map (Tzourio-Mazoyer et al., 2002), activation clusters were cluster-corrected for multiple comparisons using random field theory (z threshold > 2.3, corrected at $p < .05$). Search volumes included left amygdala (220 voxels), right amygdala (248 voxels), bilateral ACC (2713 voxels), and bilateral VMPFC (5132 voxels).

Group level analyses investigated the neural regions associated with the drift rates while controlling for starting points and differences in nondecisional time that contribute to interpreting meta-perceptions in ambiguous scenarios (in comparison with negative scenarios). Specifically, differences in drift rates ($\Delta v_{\text{Ambiguous-Negative}}$) were entered into a covariate analysis in the GLM for the contrast of Ambiguous–Negative; individual changes in starting points ($\Delta z_{\text{Ambiguous-Negative}}$) and differences in nondecisional time ($\Delta d_{\text{Ambiguous-Negative}}$) were added as nuisance regressors. For illustration purposes (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009), parameter estimates from significant activation clusters in these analyses were extracted and plotted in relation to self-flattering choices in the Ambiguous condition. Similar group level analyses investigated the neural regions associated with the starting points while controlling for individual differences in drift rates and differences in nondecisional time that contribute to interpreting meta-perceptions in ambiguous scenarios in comparison with negative scenarios.

RESULTS

Behavioral Results

A repeated-measures ANOVA tested whether Condition had an effect on absolute value of drift rates, on starting points, and on differences in nondecisional time, which shape meta-perceptions. As hypothesized, there was a main effect of Condition on drift rates (i.e., rates of evidence accumulation: $F(2, 106) = 3.69$, $p < .05$). On average, drift rates in the Ambiguous condition (mean = 0.53, $SD = 0.39$) were not significantly different than those in the Negative condition (mean = 0.63, $SD = 0.41$; $t(53) = 1.35$, $p > .05$) but were significantly slower

than drift rates in the Positive condition (mean = 0.75, SD = 0.51; $t(53) = 1.47$, $p < .05$). Furthermore, the extent to which participants had faster drift rates in the Ambiguous condition than in the Negative condition predicted the extent to which they interpreted the Ambiguous condition in a self-flattering matter ($r = .35, p < .05$). (Although the drift rates are partially estimated from the responses, this analysis demonstrates that there is a positive rather than a negative relation to self-flattering meta-perceptions). In other words, the Ambiguous condition, on average, reflected the slower evidence accumulation process that would be expected if the larger number of plausible interpretations was considered. Furthermore, the extent to which participants chose self-flattering interpretations in the Ambiguous condition was associated with their tendency to require less evidence accumulation in the Ambiguous condition (than in the Negative condition).

The other parameter estimates also reflected hypothesized effects. There was a main effect of Condition on starting points ($F(2, 106) = 6.62, p < .05$). Starting points in the Ambiguous condition (mean = 0.49, SD = 0.11) were significantly lower than starting points in the Positive condition (mean = 0.56, SD = 0.14; $t(53) = 3.10$, $p < .05$), but not significantly different than starting points in the Negative condition (mean = 0.48, SD = 0.13; $t(53) = -0.73$, $p > .05$). That is, participants did not show a significant expectation toward the flattering or unflattering interpretation in the Ambiguous and Negative conditions as their starting points were near the midpoint between the two choices. There was no main effect of Condition on differences in nondecisional time ($F(2, 106) = 0.62, p > .05$; Ambiguous condition: mean = 0.01, SD = 0.10; Positive condition: mean = 0.00, SD = 0.11; Negative condition: mean = −0.02, SD = 0.13).

Model fit as assessed by the KS statistic indicated that the diffusion model adequately accounted for the empirical data (mean $p = .89$, $SD = .12$). In fact, none of the participants showed a significant deviation ($p < .05$) of the predicted values from the empirical values.

In addition, consistent with the pilot testing of the stimuli, repeated-measures ANOVAs found significant Condition effects on response times ($F(2, 106) = 23.93$, $p < .001$) and ratings ($F(2, 106) = 314.60, p < .001$). Participants took the longest to decide on interpretations of the ambiguous scenarios, and their interpretations, on average, fell in between the positive and negative scenarios. RTs in the Ambiguous condition (mean = 1934.22, SD = 468.78) were slower than RTs in the Positive condition (mean = 1807.48, SD = 411.01; $t(53) = 7.49$, $p <$ .05) and in the Negative condition (mean = 1873.39, SD = 447.35; $t(53) = 3.86$, $p < .05$). Participants were faster to respond in the Positive condition compared with the Negative condition ($t(53) = 3.03$, $p < .05$). Participants were more likely to choose the self-flattering interpretations in the Positive condition (mean = 0.78, SD = 0.11) than in the Negative condition (mean = 0.32, SD = 0.16; $t(53) = 21.19$, $p < .05$). Self-flattering interpretations were chosen less often for the Ambiguous condition (mean = 0.60, SD = 0.14) than the Positive condition ($t(53) = -11.85, p < .05$) but more often than the Negative condition ($t(53) = 15.51, p < .05$).

Imaging Results: VMPFC Activity Is Associated with Relatively Faster Drift Rates for Interpreting Meta-perceptions in Ambiguous Scenarios

Increased VMPFC (BA 11) activation predicted the extent to which drift rates for meta-perceptions in the Ambiguous condition were faster than those in the Negative

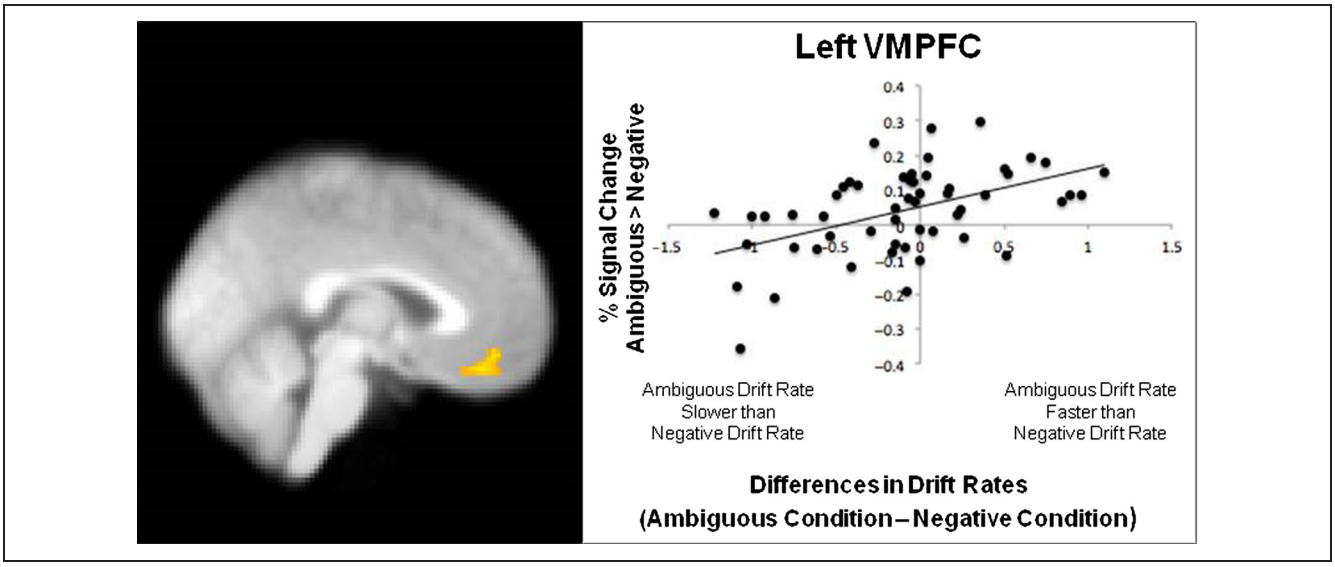

Figure 2. Neural regions associated with relatively faster drift rates for meta-perceptions in ambiguous scenarios (compared with negative scenarios). Left VMPFC cluster (BA 11, peak = −4, 38, −18; $z = 3.58$, $k = 144$) activity is positively related to the extent to which participants required less evidence accumulation for a meta-perceptive interpretation. This plot is for illustrative purposes (i.e., Kriegeskorte et al., 2009).

condition (Figure 2). In other words, VMPFC activation increased to the extent that participants required less evidence accumulation in the Ambiguous condition (compared with the Negative condition). Furthermore, an exploratory analysis was consistent with the association between requiring shallower evidence accumulation in the Ambiguous condition and a self-flattering interpretation; the increased VMPFC activation was marginally associated with choosing a self-flattering interpretation in the Ambiguous condition ($r = .25, p = .07$).

None of the other ROIs showed a significant association with the change in drift rate across the Ambiguous and Negative conditions or an association with changes in starting points across conditions. There were also no significant effects in the ROIs for the comparison of the Ambiguous and Positive conditions.

DISCUSSION

The current study is the first study to use a model-based approach to characterize the psychological significance of neural associations with meta-perceptions and the first to examine motivated meta-perceptions. Activation in the VMPFC (BA 11), a region previously associated with motivated self-perception (for reviews, see Delgado et al., 2016; Flagan & Beer, 2013) and meta-perception (Veroude et al., 2014; Pfeifer et al., 2009; Ochsner et al., 2005), was significantly associated with evidence accumulation processes for meta-perceptions in relatively ambiguous conditions compared with negative conditions. The Ambiguous condition consisted of scenarios for which there was no social consensus on their interpretation; however, it is worth noting that the findings in the current study may reflect ambiguity arising from a lack of social consensus or from uncertainty as participants took longest to respond in the Ambiguous condition. Furthermore, VMPFC activation showed an association with self-flattering meta-perceptions. Taken together, the present findings shed new light on the precise functions of VMPFC in social evaluation and raise new research avenues to further refine our understanding of how the brain supports the interpretation of social situations.

The current findings expand neural investigations of social cognition by investigating meta-perceptions using a model-based approach. Although much less attention has been paid to the neural associations of meta-perceptions than self- or other-evaluations, previous studies suggested that regions of VMPFC might be involved (e.g., Veroude et al., 2014; Pfeifer et al., 2009; Ochsner et al., 2005). The model-based approach of the current study made it possible to build on previous studies by pinpointing whether the VMPFC is associated with evidence accumulation processes, a priori expectations (which may be imposed on interpretation regardless of available information), or both. Traditional self-report and RT measures make it difficult to distinguish between these various explanations. Participants may not know how they reached an interpretation (e.g., Nisbett & Wilson, 1977), and both selective evidence accumulation and the imposition of a priori expectations without examination of available evidence would cause RTs to be faster. The application of DDM showed that the VMPFC may play a role in the evidence accumulation process that shapes meta-perceptions. VMPFC may have mediated the different rate at which participants gathered evidence when first presented with an ambiguous social scenario. It is also possible that VMPFC predicts a faster evidence accumulation rate at the point of the decision prompt (rather than during the scenario presentation) as the scenario and decision prompts were not jittered apart. Future research should investigate whether VMPFC mediates evidence accumulation when a situation is first processed or at the point at which a commitment to an interpretation is made.

The current findings point to the need to understand whether the VMPFC also plays a similar evidence accumulation role in self- versus other-evaluation. If this were the case, it is possible that the robust VMPFC activation associated with self-evaluations (in comparison with evaluations of others: for a review, see Denny, Kober, Wager, & Ochsner, 2012) reflects a preferential evidence accumulation process. Such a finding would be informative as most accounts of the role of VMPFC in self-evaluation focus on positive affect (see Roy, Shohamy, & Wager, 2012, for a review, and Chavez et al., 2016). Although it may be that self-evaluation and positive affect share a common valence or experiential feeling, which is mediated by the VMPFC, the application of the model-based approach in the current study makes it possible to test another possibility: similar facilitation of evidence accumulation. This possibility would be consistent with behavioral research, which has found that both self-evaluation and positive affect are associated with relatively greater processing fluency (e.g., Winkielman & Cacioppo, 2001; Klein & Kihlstrom, 1986).

The current findings also build on existing research, which has found a consistent role of VMPFC in self-evaluations that are self-protective in nature (for reviews, see Delgado et al., 2016; Flagan & Beer, 2013). The current research is the first to suggest that VMPFC may also support self-protective processing through self-flattering interpretations of others' thoughts about the self. Future research is needed to understand whether VMPFC plays a role in self-protective processing through evidence accumulation from internal thoughts, external information, or some combination of the two. For example, in the domain of social cognition, a VMPFC functional network is theorized to support bottom–up reward processing and is characterized in contrast to a dorsomedial pFC functional network that is theorized to support metacognitive processes. A dichotomous VMPFC and dorsomedial pFC framework suggests that VMPFC supports self-protective processing of meta-perceptions by drawing on internally generated rewarding thoughts (as may be the case in contexts involving self-evaluation), interacting with the

metacognitive functional network in the particular case of meta-perceptions, or both.

In summary, the current study used a model-based approach to more precisely characterize the role of VMPC in meta-perceptions particularly in conditions where their self-flattering nature is not evident. The study found that the VMPFC mediates evidence accumulation processes in meta-perceptions that may be used to arrive at self-flattering interpretations and suggested that the VMPFC's role in self-protective social cognition may extend to self-flattering meta-perceptions. As this study demonstrates, adaptation of model-based approaches from perceptual research may advance our understanding of the precise functions carried out by neural regions known to be involved in social cognition.

Acknowledgments

This work was supported by the National Science Foundation (NSF-BCS-1147776 to J. S. B. and DGE-1110007 to T. F.).

Reprint requests should be sent to Jennifer S. Beer, Department of Psychology, The University of Texas at Austin, 1 University Station A8000, Austin, TX 78712, or via e-mail: beerutexas@gmail.com.

REFERENCES

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20,* 201–248.

Beard, C., & Amir, N. (2009). Interpretation in social anxiety: When meaning precedes ambiguity. *Cognitive Therapy and Research, 33,* 406–415.

Beer, J. S. (2007). The default self: Feeling good or being right? *Trends in Cognitive Sciences, 11,* 187–189.

Beer, J. S., & Hughes, B. L. (2010). Neural systems of social comparison and the "above-average" effect. *Neuroimage, 49,* 2671–2679.

Bengtsson, S. L., Dolan, R. J., & Passingham, R. E. (2011). Priming for self-esteem influences the monitoring of one's own performance. *Social Cognitive and Affective Neuroscience, 6,* 417–425.

Chavez, R. S., Heatherton, T. F., & Wagner, D. D. (2016). Neural population decoding reveals the intrinsic positivity of the self. *Cerebral Cortex*. doi:10.1093/cercor/bhw302.

Cunningham, W. A., Van Bavel, J. J., & Johnsen, I. R. (2008). Affective flexibility: Evaluative processing goals shape amygdala activity. *Psychological Science, 19,* 152–160.

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

D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., et al. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience, 19,* 935–944.

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage, 19,* 430–441.

Delgado, M. R., Beer, J. S., Fellows, L. K., Huettal, S. A., Platt, M. L., Quirk, G. J., et al. (2016). Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nature Neuroscience, 19,* 1545–1552.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 24,* 1742–1752.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63,* 568–584.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Flagan, T., & Beer, J. S. (2013). Three ways in which midline regions contribute to self-evaluation. *Frontiers Human Neuroscience, 7,* 450.

Germar, M., Albrecht, T., Voss, A., & Mojzisch, A. (2016). Social conformity is due to biased stimulus processing: Electrophysiological and diffusion analyses. *Social Cognitive and Affective Neuroscience, 11,* 1449–1459.

Hughes, B. L., & Beer, J. S. (2012). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *Neuroimage, 61,* 889–898.

Hughes, B. L., & Beer, J. S. (2013). Protecting the self: The effect of social-evaluative threat on neural representations of self. *Journal of Cognitive Neuroscience, 25,* 613–622.

Jenkins, A. C., & Mitchell, J. P. (2010). Mentalizing under uncertainty: Dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex, 20,* 404–410.

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience, 6,* 211–218.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17,* 825–841.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self?: An event-related fMRI study. *Journal of Cognitive Neuroscience, 14,* 785–794.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. *NeuroReport, 14,* 2317–2322.

Klein, S. B., & Kihlstrom, J. F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology: General, 115,* 26–38.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology, 3,* 193.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12,* 535–540.

Markus, H., Smith, J., & Moreland, R. L. (1985). Role of the self-concept in the perception of others. *Journal of Personality and Social Psychology, 49,* 1494.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making, 5,* 437–449.

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., et al. (2010). Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine, 63,* 1144–1153.

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., & Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of Cognitive Neuroscience, 18,* 1586–1594.
Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage, 59,* 2636–2643.
Neta, M., Kelley, W. M., & Whalen, P. J. (2013). Neural responses to ambiguity involve domain-general and domain-specific emotion processing systems. *Journal of Cognitive Neuroscience, 25,* 547–557.
Neta, M., Norris, C. J., & Whalen, P. J. (2009). Corrugator muscle responses are associated with individual differences in positivity-negativity bias. *Emotion, 9,* 640–648.
Neta, M., & Whalen, P. J. (2010). The primacy of negative interpretations when resolving the valence of ambiguous facial expressions. *Psychological Science, 21,* 901–907.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231.
Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D. E., Kihlstrom, J. F., et al. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage, 28,* 797–814.
Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development, 80,* 1016–1038.
Preuss, G. S., & Alicke, M. D. (2009). Everybody loves me: Self-evaluations and metaperceptions of dating popularity. *Personality and Social Psychology Bulletin, 35,* 937–950.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59.
Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences, 16,* 147–156.
Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. A. (2009). A neural mechanism of first impressions. *Nature Neuroscience, 12,* 508–514.
Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42,* 9–34.
Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3,* 102–116.
Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17,* 143–155.
Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23,* S208–S219.
Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–210.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.
Veroude, K., Jolles, J., Croiset, G., & Krabbendam, L. (2014). Sex differences in the neural bases of social appraisals. *Social Cognitive and Affective Neuroscience, 9,* 513–519.
Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology, 44,* 1048–1056.
Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32,* 1206–1220.
Voss, A., & Schwieren, C. (2015). The dynamics of motivated perception: Effects of control and status on the perception of ambivalent stimuli. *Cognition and Emotion, 29,* 1411–1423.
Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39,* 767–775.
White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 385–398.
Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology, 81,* 989–1000.