Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition

Brent L. Hughes, Jennifer S. Beer *

Department of Psychology and Imaging Research Center, The University of Texas at Austin, Austin, TX 78712, USA

**ARTICLE INFO**

*Article history:*  
Accepted 5 March 2012  
Available online 13 March 2012

**Keywords:**  
Social cognition  
Self-regulation  
Motivation  
Emotion  
fMRI

**ABSTRACT**

Recent research has begun to identify neural regions associated with self-serving cognition, that is, the tendency to make claims that cast the self in an overly flattering light, yet little is known about the mechanisms supported by neural activation underlying self-serving cognition. One possibility suggested by current research is that MOFC, a region that shows reduced recruitment in relation to self-serving cognition, may support changes in the decision thresholds that influence whether information should be expressed in an evaluation. The current fMRI study addresses this question by combining a signal detection approach and a contextual manipulation that permits the measurement of changes in decision threshold. Participants evaluated their familiarity with blocks of existent and nonexistent information when they believed that self-serving claims of knowledge could either be exposed (accountable condition) or not (unaccountable condition). When held accountable, participants tended to shift their decision thresholds in a conservative (i.e., less self-serving) direction and showed greater activation in orbitofrontal cortex (OFC), medial prefrontal cortex (MPFC) and dorsal anterior cingulate cortex (dACC). Furthermore, the extent to which participants adopted more conservative (i.e., less self-serving) decision thresholds as a function of context (i.e., accountability), the more they recruited MOFC activation. These findings refine current knowledge about the mechanisms performed by neural regions involved in self-serving cognition and suggest a role for MOFC in changing decision thresholds that influence whether information should be expressed in an evaluation.

© 2012 Elsevier Inc. All rights reserved.

**Introduction**

Recent research has begun to identify neural regions associated with self-serving cognition, that is, the tendency to make claims that cast the self in a favorable light, but the psychological mechanisms of this association are far from understood (Beer, 2007). The motivation to cast the self in a flattering light is one of the most pervasive motivational influences on social cognition (Sedikides and Gregg, 2008). Behavioral research has illustrated a number of ways in which people are self-serving in their judgments and these phenomena have recently caught the attention of researchers interested in bridging the gap between the psychological and neural literatures on social cognition. A recent wave of fMRI, lesion, and ERP research has investigated overly confident evaluations of success on a trivia task (Beer et al., 2010), exaggerated desirability of one’s own personality in comparison to the personality of peers (Beer and Hughes, 2010; Hughes and Beer, in press), self-perceptions of task performance that are more favorable than the perceptions of trained judges (Beer et al., 2006), self-serving attributions for hypothetical social events and task performance (Blackwood et al., 2003; Krusemark et al., 2008), and memory for positive compared to negative feedback (Somerville et al., 2010). Whereas conditions associated with self-serving evaluations are associated with the reduced recruitment of orbitofrontal cortex (OFC), medial prefrontal cortex (MPFC), and dorsal anterior cingulate (dACC) across a wide variety of domains (Beer and Hughes, 2010; Beer et al., 2010; Blackwood et al., 2003; Hughes and Beer, in press; Krusemark et al., 2008), only the medial OFC is consistently negatively modulated by individual differences in self-serving evaluations (Beer and Hughes, 2010; Beer et al., 2010; Hughes and Beer, 2011; Somerville et al., 2010) and, in fact, damage to OFC is associated with self-serving evaluation (Beer et al., 2006). For example, reductions in MOFC activity predict sample-level parametric increases in confidence for incorrect task performance (Beer et al., 2010) and exaggerations of the desirability of one’s own personality (and the personality of intimate others when compared to other people: Beer and Hughes, 2010; Hughes and Beer, in press). Despite the robust association between reduced MOFC activation and self-serving cognition across a variety of paradigms, the question of why MOFC activation attenuates self-serving cognition remains undressed. To address this question, the present study refines current knowledge about the mechanisms supported by MOFC in self-serving cognition by examining the possible relation between MOFC activation and one mechanism suggested by previous research: shifts in decision threshold.
One possible implication of previous research is that MOFC is associated with shifts in self-serving responses because it mediates shifts in the decision thresholds that influence the extent to which accessible internal information is expressed in evaluations. For example, previous research has speculated that the relation between MOFC activation and reduced overconfidence in task performance might reflect shifting thresholds for allowing perceived task familiarity to influence confidence judgments (Beer et al., 2010). MOFC activation attenuated self-serving patterns by marking conservative shifts in confidence estimates (i.e., significant negative parametric modulation across the study sample) about the veracity of answers to trivia questions in conditions where answers were indeed incorrect (Beer et al., 2010). Behavioral research using the same paradigm finds that overconfidence in performance arises when judgments liberally rely on gut intuitions about perceived familiarity with the trivia domain (rather than actual knowledge for a particular question, etc.: Klayman et al., 1999). Therefore, it is possible that parametric increases in OFC activation were associated with less confidence in wrong answers as the sample became more and more conservative about allowing generalized familiarity to influence judgments of confidence. A similar explanation could also account for the relation between MOFC activation and reductions in other kinds of self-serving judgments, such as people’s tendency to claim that they have more desirable personality traits and fewer undesirable personality traits than their peers (Chambert and Windschitl, 2004). People’s self-serving tendency to claim that they (or their romantic partners) have more desirable personalities than their peers is attenuated when the comparisons are made for narrowly-constrained (versus broadly-constrained) personality traits (Dunning et al., 1989); attenuation as a function of trait breadth is predicted by MOFC activation (Beer and Hughes, 2010; Hughes and Beer, in press). The increased MOFC activation may reflect a shift in decision threshold because one potential effect of trait breadth manipulations is that they shift how liberally one can construe a trait as self-relevant and, consequently, judge oneself to be better than one’s peers on that trait (Dunning et al., 1989). For example, talent is a relatively broadly-constrained trait (Hampson et al., 1986). There are many different kinds of behavior that a person could use to idiosyncratically define talent in order to boost how talented they are in comparison to peers. However, the threshold for calling oneself especially talented is relatively more conservative than for talent because far fewer behaviors actually relate to tidiness (making idiosyncratic definitions more difficult and reducing the self-serving nature of comparisons). Therefore, the increased MOFC activation may reflect the extent to which participants were sensitive to the more conservative decision threshold inherent in claiming narrowly-constrained traits as especially self-relevant. Although a shift in the relative liberality of decision threshold may account for OFC modulation in previous studies of self-serving responses, the paradigms used in previous neural studies were not designed to directly measure changes in decision threshold. A more rigorous test of the mechanism implied by previous neural findings requires research that directly measures changes in decision threshold in relation to the expression of self-serving responses.

Previous behavioral research provides clear methods to measure decision thresholds underlying self-serving claims and the conditions in which those thresholds are shifted. Specifically, probing participants’ familiarity with scholarly items while manipulating the accountability of their claims has been shown to be a fruitful method for examining shifts in the decision thresholds associated with self-serving responses (Paulhus et al., 2003). Accountability refers to the expectation of having to justify one’s judgments, such as having to elaborate on the reasons for judgment or having the judgment evaluated by a third-party (Lerner and Tetlock, 1999; Sedikides et al., 2002). This expectation may be explicit (Sedikides et al., 2002) or implicit (Tice et al., 1995). Previous behavioral research has shown that both explicit and implicit accountabilities reduce the expression of self-serving claims in various domains, such as claiming familiarity with scholarly knowledge (Paulhus et al., 2003), overly confident evaluations of success (Tetlock and Kim, 1987; Tice et al., 1995), exaggerated desirability of one’s personality in comparison to the personality of peers (McKenna and Myers, 1997), and self-perceptions of task performance that are more favorable than the perceptions of other people (Sedikides et al., 2002). Therefore, accountability is a robust context that affects the magnitude of self-serving responses. Furthermore, previous research on accountability and overclaiming of knowledge (i.e., self-serving responses) provides an approach for measuring shifts in decision threshold. As mentioned above, people tend to claim knowledge of scholarly concepts beyond what they could possibly know; this tendency is reduced when participants are notified that they are rating a list of concepts that may or may not exist (i.e., any false claims will be exposed: Paulhus et al., 2003). Previous research on overclaiming has applied signal detection theory (SDT) (Green and Swets, 1966; Macmillan and Creelman, 1991) techniques to claims of knowledge. The SDT approach further reinforces the utility of overclaiming knowledge as an index of the expression of self-serving tendencies. The people who are especially likely to use liberal decision thresholds when claiming familiarity are narcissists and those motivated to deceive others into seeing them in a positive light (Bing et al., 2011; Paulhus et al., 2003; Randall and Fernandes, 1991; Tracy et al., 2009). Warning participants that information may be nonexistent reduces self-serving claims of knowledge by changing decision thresholds in a more conservative direction (Paulhus et al., 2003). This case, SDT considers a more conservative decision threshold to reflect a reduction in how liberally participants are willing to construe familiarity signals as indicative of actual knowledge. Therefore, using SDT to analyze the effects of accountability on overclaiming of knowledge makes it possible to measure changes in decision threshold that relate to self-serving responses and test its relation to MOFC activation. MOFC activation should predict the extent to which participants adopt more conservative (i.e., less self-serving) decision thresholds when self-serving claims have the potential to be exposed.

Study overview

The present functional magnetic resonance imaging (fMRI) study examined whether neural regions previously implicated in self-serving cognition relate to changes in decision thresholds underlying the extent to which judgments are self-serving. Participants evaluated their familiarity with blocks of existent and nonexistent knowledge items while being held accountable or unaccountable for their evaluations (Paulhus et al., 2003). Signal detection theory (SDT) analyses were applied to measure the shift in decision threshold across conditions (Green and Swets, 1966; Macmillan and Creelman, 1991). Based on previous behavioral research (Paulhus et al., 2003), accountability should reduce self-serving claims of knowledge by shifting the decision threshold in a conservative direction. If the association between MOFC and self-serving responses reflects its role in shifting decision thresholds, then MOFC activation should predict the extent to which decision thresholds become more conservative as a result of accountability.

Materials and method

Participants

18 right-handed participants (9 females, M age = 20.7 years, SD = 1.9 years) were recruited in compliance with the human subjects regulations of the University of Texas at Austin and compensated with $15/h or course credit. All participants were native English speakers and free from medications or psychological and/or neurological conditions that might influence the measurement of cerebral blood flow.
Behavioral paradigm

Participants completed a modified version of the over-claiming questionnaire and accountability manipulation used in previous research (Paulhus et al., 2003). To ensure they understood the task, participants completed a practice session just before entering the scanner. In the practice session, participants first received verbal instructions from the experimenter and then read through self-paced instruction screens. The instructions informed participants that they would be asked to rate their familiarity with different knowledge items (e.g., famous people, books, scientific concepts) and that these ratings would be divided into sections. The experimenter emphasized that some sections would include knowledge items that do not exist and, therefore, it would be impossible for participants to be familiar with those items. The participants learned that instructions for different sections would vary only slightly so that it was important that they read all instructions very carefully. The variance in the instructions manipulated Accountability (see below). After the instruction period, participants completed 4 blocks (2 Accountable, 2 Unaccountable) of practice trials of items not used in the scanner task. The experimenter then discussed the procedure with participants. If questions arose, participants were able to complete the practice session as many times as they needed to feel comfortable with the task.

In the scanner sessions, participants’ attention was drawn to the fact that experimenters could see their responses in real time via an interactive task to verify that the button box was registering each number from the rating scale. Participants were then presented again with the instruction screens. For the task in the scanner, participants rated their familiarity with blocks of knowledge items in two Accountability conditions (Accountable, Unaccountable; see Fig. 1). Blocks of items were preceded by task instructions (4 s) and either (1) a notice that some of the upcoming items may be nonexistent (i.e., any self-serving claims of knowledge would be exposed (Accountable blocks)) or (2) no notice (Unaccountable blocks). In order to call attention to the difference between blocks including the notice and blocks that did not include the notice, the notice was presented in a different color than all other text in the experiment (i.e., bright green compared to white for all other text). Each instruction screen was followed by a 2 sec screen with a fixation point indicating that participants should clear their minds. Participants were then presented with blocks of knowledge items that exist (e.g., Billie Holiday) or do not exist (e.g., J.D. Louis) and asked to rate their familiarity with each item. Familiarity for each item was rated on a 4-point scale from 0 (not at all familiar) to 3 (very familiar). Regardless of Accountability condition, blocks consisted of 10 items that included 6 existent and 4 nonexistent items (20 sec block; 2 sec each item). Participants completed 8 blocks (4 each for the Accountable and Unaccountable blocks) for a total of 48 existent items and 32 nonexistent items. Existent and Nonexistent items that had been normed for equivocal familiarity ratings were pseudorandomly assigned to Accountable and Unaccountable conditions for the fMRI procedure. A separate sample of 70 judges rated their familiarity with all of the items in the absence of an accountability manipulation. A random number generator was used to assign existent and nonexistent items to either the accountable or unaccountable condition. Normative familiarity did not differ across conditions (Accountable vs Unaccountable: t(69) = .35, p > .05); Accountable vs Unaccountable Nonexistent: t(69) = .58, p > .05) and did significantly distinguish between existent and nonexistent items within condition (Accountable Existent vs Nonexistent: t(69) = 11.61, p < .05); Unaccountable Existent vs Nonexistent: t(69) = 10.99, p < .05; see Supplemental Table 1). Further, two pilot studies that did not employ the pseudorandom procedure used in the fMRI study found convergent behavioral results with the fMRI study (see Supplemental Table 2).

fMRI data were collected in one 7 minute, 6 second run consisting of pseudorandomized Accountable, Unaccountable, and “Catch” blocks. “Catch” blocks were included to establish that neural differences between Accountable and Unaccountable blocks were not merely due to the presence of the Accountability cue screen (see Fig. 1). The Accountable “Catch” blocks consisted of (a) the 4 sec Accountable cue screen, (b) a 2 sec fixation point, and (c) a 20 sec screen that instructed participants to clear their minds. The Unaccountable “Catch” blocks consisted of (a) the 4 sec Unaccountable cue screen, (b) a 2 sec fixation

Fig. 1. Stimuli and timing in over-claiming bias task. Participants saw an instruction screen that either did or did not provide an Accountability cue (i.e., some of the items may be nonexistent, portrayed in black and white but in the experiment this phrase was presented in green while all other text was presented in white), followed by a fixation cross, followed by a block of 10 trials in which participants rated their familiarity with existent (e.g., Billie Holiday) and nonexistent (e.g., J.D. Louis) knowledge items. The Accountable “Catch” blocks and Unaccountable “Catch” blocks were the same as their respective experimental blocks except that the phrase “Please Clear your Mind” was substituted for the knowledge item probes.
point, and (c) a 20 sec screen that instructed participants to clear their minds. Stimuli were projected onto a screen mounted on the bed of the scanner and head motion was limited using foam padding. E-prime on a Windows XP was used to present stimuli and collect responses.

**Behavioral indices**

Conceptually, the goal of creating behavioral indices was to identify an index of shift in decision threshold across conditions that marked changes in self-serving responses. This index is not a measure of task performance, in other words, it was not just how good participants were at identifying whether items existed or not as a function of accountability. Following previous behavioral research on the effect of accountability on over-claiming (e.g., Paulhus et al., 2003), signal detection theory (SDT: Green and Swets, 1966) was used to model two indices: shifts in participants’ tendency to make self-serving claims of knowledge (shifts in decision threshold (c)) and shifts in their ability to discriminate between existent and nonexistent items (shifts in discriminability (d')). The decision threshold (c) provides a measure of self-serving cognition because it is theorized to reflect how strong the sense of familiarity with items is needed in order to claim knowledge. For example, a participant may require very strong evidence of familiarity in order to claim knowledge and therefore may over-claim less than a self-serving participant that requires much weaker evidence of familiarity in order to claim knowledge (see Fig. 2A). Participants’ tendency to shift decision thresholds can be measured by contrasting the thresholds used across accountability conditions. Discriminability (d'), on the other hand, represents participants’ ability to distinguish items that do exist from items that do not exist rather than their tendency to make self-serving claims (see Fig. 2B). Discriminability (d') does not necessarily reflect the self-serving nature of knowledge claims because it does not provide information about the strategy for claiming knowledge, how it changes across contexts, or the strength of the evidence needed in order to claim knowledge. For example, two participants might be quite good at telling which items exist and which do not (both have high discriminability) but the participants will differ in how self-serving their claims are if they differ in how much they claim to know things for which they only have very weak feelings of familiarity (high versus low thresholds: see Fig. 2A).

Several steps were completed to compute shifts in decision threshold (c) and discriminability (d'). Responses were classified into: (1) hits: claims that existent items are familiar; (2) false alarms: claims that nonexistent items are familiar; (3) misses: claims that nonexistent items are not familiar; and (4) correct rejections: claims that nonexistent items are not familiar. Following previous behavioral research on the effect of accountability on over-claiming (e.g., Paulhus et al., 2003), the decision threshold (c) and discriminability (d') were calculated at each of the 3 cutoffs on the 0–3 rating scale (i.e., cutoff of 0, 1, and 2) (also see Macmillan and Creelman, 1991). For the 0 cutoff, responses greater than 0 were classified as a hit or false alarm, and responses of 0 were classified as a miss or correct reject. The same process was repeated for the cutoff of 1 and 2. The hit rate (HR) was the proportion of the 48 existent items on which participants gave a familiarity rating greater than the cutoff. The false-alarm rate (FAR) was the proportion of the 32 nonexistent items on which participants gave a familiarity rating greater than the cutoff. Decision threshold (c) and discriminability (d') were then calculated at each cutoff for each of the Accountable and Unaccountable conditions, and averaged to get a final value of decision threshold (c) and discriminability (d') for each participant for each of the Accountable and Unaccountable conditions (Macmillan and Creelman, 1991). Shifts in threshold and discriminability were measured by subtracting each measure in the Unaccountable condition from its respective measure in the Accountable condition. Each of these measures is described below in greater detail.

From an SDT perspective, the decision threshold (c) provides information about the self-serving nature of knowledge claims because it reflects the strength of the evidence needed in order to claim knowledge.

**Fig. 2.** Examples of familiarity distributions and decision thresholds for existent and nonexistent items. (A) Vertical lines illustrate conservative and liberal decision thresholds. Thresholds become more conservative (higher) as they move toward the strong end of the distribution of internal signals of familiarity (only the items that generate internal familiarity signals that are stronger than the conservative threshold will get claimed). Thresholds become more liberal as they move downward toward the ‘weak’ end of the distribution (much weaker internal familiarity signal is needed to claim knowledge as compared to the more conservative threshold on the left). (B) Degree of overlap between distributions for nonexistent and existent items illustrates low and high discriminability. High degree of overlap indicates low discriminability (d') (internal familiarity signals do not do much to distinguish nonexistent items from existent items) whereas smaller degree of overlap indicates high discriminability (d') (on average, internal familiarity signal is stronger for existent compared to nonexistent items).
knowledge with an item (Macmillan and Creelman, 1991; Paulhus et al., 2003). An observer with a high decision threshold will have low hit rates and false alarm rates, whereas an observer with a low decision threshold will have high hit rates and false alarm rates:

\[ d = \frac{z(HR) - z(FAR)}{2} \] (1)

On the other hand, SDT takes the perspective that discriminability \((d')\) reflects the ability to discriminate between existent and nonexistent items, rather than the self-serving nature of knowledge claims. An observer with high discriminability will have a high hit rate and a low false alarm rate, whereas an observer with low discriminability will have more similar hit rates and false alarm rates. According to SDT, an observer experiences an internal sense of familiarity to an item, and this internal sense of familiarity represents a point on a continuum of familiarity. Existent items (e.g., Billie Holiday) and nonexistent items (e.g., J.D. Louis) are theorized to generate a stronger sense of familiarity than nonexistent items because it is not known (i.e., it may generate a weak internal familiarity signal), and an observer may sometimes feel somewhat familiar with a nonexistent item because it is similar to something they know (i.e., it may generate a strong internal familiarity signal). Therefore, the most commonly applied SDT model assumes that the internal familiarity signals generated by these existent and nonexistent items are normally distributed and overlap with each other (Wickens, 2002; see Fig. 2). The distance between the means of the existent and nonexistent distributions represents the discriminability between existent and nonexistent items (HR = hit rate; FAR = false alarm rate; Macmillan and Creelman, 1991):

\[ d = \frac{z(HR) - z(FAR)}{2} \] (2)

**MRI data acquisition**

All images were collected on a 3.0-T GE Signa EXCITE scanner at the University of Texas at Austin Imaging Research Center (UT-IRC). Functional images were acquired with a GRAPPA sequence (TR = 2000 ms, TE = 30 ms, FOV = 240, voxel size 2.5 x 2.5 x 3.3 mm) with each volume consisting of 35 axial slices oriented to the AC–PC line. These parameters were implemented to optimize coverage of the orbitofrontal cortex without sacrificing whole-brain acquisition (e.g., Clayton et al., 2006, see Mehta and Beer, 2009 for example coverage and signal from the UT-IRC, see Supplementary Fig. 1 for coverage and SNR from the current study). A high resolution SPGR T1-weighted image was also acquired from each subject.

**MRI data analysis**

Statistical analyses were conducted using SPM2 (Wellcome Department of Cognitive Neurology). Functional images were reconstructed from k-space using a linear time-interpolation algorithm to double the effective sampling rate. Volume images were corrected for slice-timing skew using temporal sinc-interpolation and for movement using rigid-body transformation parameters. Functional data and structural data were co-registered and normalized into a standard anatomical space (2 mm isotropic voxels) based on the EPI and T1 templates (Montreal Neurological Institute), respectively. Images were smoothed with an 8-mm FWHM Gaussian kernel. A high-pass filter with a cutoff of 128 s was applied to remove within-session drifts.

At the individual subject level, a fixed-effects analysis modeled the Accountable blocks, Unaccountable blocks, and the Accountable and Unaccountable “Catch” blocks using a canonical block hemodynamic response function. A general linear model analysis created contrast images for each participant. Contrasts were calculated to examine neural activation in the contrasts of the Accountable block > Unaccountable block, Unaccountable block > Accountable block, and Accountable “Catch” block > Unaccountable “Catch” block. At the group level, contrasts from each participant were used in a second-level analysis treating participants as a random effect. The group average SPM(t) maps were corrected for multiple comparisons at the cluster level (based on the CorrClusTh algorithm created by Thomas Nichols: http://www.sph.umich.edu/~nichols/JohnsGems5.html) in hypothesized neuroanatomical regions (MPFC, MOFC, LOFC, and dACC, defined by the Automated Anatomical Labeling map: Tzourio-Mazoyer et al., 2002). For all results, the threshold was set to a minimum of 165 contiguous voxels at a voxel-wise threshold of p < .005, to achieve a statistical threshold of p < .05, corrected for multiple comparisons at the cluster level.

Correlation analyses tested whether individual differences in shifts in decision threshold \((c)\) and discriminability \((d')\) modulated neural activation identified in the main contrasts. Parameter estimates from significant clusters identified in the main contrasts were extracted using Marsbar (Brett et al., 2002). The parameter estimates from significant clusters were tested for significant correlation with individual differences in each behavioral index (controlling for the influence of the other behavioral index; shift in decision threshold \((c)\) and discriminability \((d')\)) were marginally correlated \((r = .43, p = .07)\). First, parameter estimates were tested for significant correlation with individual differences in Accountability’s effect on shifts in decision threshold (Accountable threshold \((c)\) – Unaccountable threshold \((c)\); controlling for shifts in discriminability \((d')\)). The difference in decision thresholds between the Accountable versus Unaccountable blocks indexes the degree to which participants’ decision thresholds became more conservative as a result of Accountability. Greater values indicate more conservative decision thresholds, or reduced over-claiming, as a function of Accountability. From an SDT perspective, participants with larger threshold difference values make fewer self-serving claims of knowledge after being held Accountable because much stronger evidence of familiarity is needed in order to claim knowledge.

Second, parameter estimates were tested for significant correlation with individual differences in Accountability’s effect on shifts in discriminability (Accountable discriminability \((d')\) – Unaccountable discriminability \((d')\); controlling for shifts in decision threshold \((c)\)). The difference in discriminability between the Accountable versus Unaccountable blocks indexes the degree to which Accountability affected the ability to discriminate existent from nonexistent items. Greater difference values indicate increased discriminability between existent and nonexistent items as a function of Accountability. From an SDT perspective, participants with larger discriminability difference values are better able to discriminate between existent and nonexistent items after being held Accountable because they have a stronger sense of familiarity with existent items than nonexistent items. Finally, analyses involving parameter estimates from MPFC were conducted using the indices described above but used robust regression. One participant had outlying parameter estimate values in MPFC (i.e., more than 3 standard deviations away from the mean). Therefore, all individual difference analyses with MPFC used robust regression to down-weight the influence of the outlier.

**Results**

**Task performance**

Decision thresholds became significantly more conservative and discriminability was significantly reduced as a function of accountability. Consistent with previous research (Paulhus et al., 2003), people made fewer self-serving claims of knowledge after receiving an
Accountability cue (see Table 1). Specifically, participants’ decision threshold (c) was significantly more conservative in the Accountable compared to the Unaccountable blocks (Accountable over-claiming: M = .91, SD = .25; Unaccountable over-claiming: M = .79, SD = .28; t(17) = 2.46, p < .05). That is, participants’ decision threshold for claiming knowledge became more conservative (i.e., less self-serving) after being cued that self-serving claims of knowledge would be exposed.

In addition, discriminability (d′) was significantly reduced in the Accountable versus Unaccountable blocks (t(17) = 2.60, p < .05). Although discriminability (d′) was reduced by Accountability, participants were able to differentiate existent from nonexistent items in both the Accountable (d′ M = 1.07, SD = .52) and Unaccountable (d′ M = 1.44, SD = .54) blocks. Specifically, discriminability (d′) in the Accountable and Unaccountable blocks was significantly different from 0, the point at which there is no discriminability between existent and nonexistent items (Accountable: t(17) = 8.79, p < .05; Unaccountable: t(17) = 11.39, p < .05). Additionally, participants were not merely guessing along the 4-point rating scale. Hit rates were above chance-level (.25) in the Accountable (M = .38, SD = .14, t(17) = 3.81, p < .05) and Unaccountable (M = .48, SD = .16, t(17) = 5.81, p < .05) blocks. Further, participants’ raw familiarity ratings on average were significantly higher for existent items compared to nonexistent items in both the Accountable (existent items: M = 1.13, SD = .40; nonexistent items: M = .23, SD = .19, t(17) = 8.65, p < .05) and Unaccountable (existent items: M = 1.45, SD = .48; nonexistent items: M = .18, SD = .17, t(17) = 10.93, p < .05) conditions.

Finally, the effect of Accountability on the decision threshold (c) and discriminability (d′) was not merely the result of time spent on the task. No significant differences were found in reaction times for the Accountable versus Unaccountable blocks (Accountable RT: M = 1.22 s, SD = .12 s; Unaccountable RT: M = 1.18 s, SD = .10 s; t(17) = 1.34, ns).

**Imaging results**

**Conditions that reduce self-serving responses are associated with increased OFC, MPFC, and dACC activation**

Consistent with previous research on self-serving cognition (Beer and Hughes, 2010; Beer et al., 2010; Blackwood et al., 2003; Hughes and Beer, in press; Krusemark et al., 2008), the current study on self-serving claims of knowledge found that the main effect of Accountability was associated with increased activation in medial OFC (BA 11 peak = −6, 58, −20), lateral OFC (right BA 47 peak = −32, 44, −18; left BA 47 peak = −30, 56, −12), MPFC (BA 10 peak = −16, 64, −2; BA 9 peak = −2, 34, 62), and dACC (BA 24 peak = −4, 14, 34) (see Table 2, Fig. 3). Activity in these regions cannot be accounted for by reaction to the presentation of the Accountability cue in the Accountability blocks. The Accountable “Catch” block versus Unaccountable “Catch” block contrast did not identify neural activation clusters in any of the a priori ROIs (see Table 3 for an additional whole brain analyses contrasting Accountable “Catch” blocks with Unaccountable “Catch” blocks). No significant activation clusters were found for the Unaccountable > Accountable blocks.

*Individual differences in shifts in decision threshold but not shifts in discriminability modulate MOFC*

Also consistent with previous research, all of the activation clusters found in a priori regions-of-interest for the main contrasts, only MOFC activation showed a significant positive association with shifts in decision threshold (c) as a function of Accountability (see Fig. 4) and none showed significant associations with shifts in discriminability (d′) as a function of Accountability. Specifically, activation in the MOFC region identified in the Accountable-Unaccountable contrast predicted the extent to which participants adopted a more conservative (i.e., less self-serving) decision threshold as a function of Accountability (r = 0.52, p < .05, controlling for discriminability; see Fig. 4). No significant relation was found between MOFC activation and shifts in discriminability as a function of Accountability r = −.31, ns, controlling for decision threshold; see Fig. 4) nor decision threshold (c) or discriminability (d′) within the Accountable or Unaccountable condition (rs range from .15 to −.37, ns). In other words, MOFC activation counteracted self-serving responses by significantly predicting the extent to which participants shifted their decision threshold in a conservative manner across contexts and was not simply related to the extent to which they withheld ratings of familiarity or discriminated existent items within a context.

**Discussion**

The present study provides the first evidence for a decision threshold shift function of the MOFC activation associated with self-serving cognition. Current research on self-serving cognition has found robust evidence for a negative association between OFC function and self-serving cognition across a variety of domains (Beer and Hughes, 2010; Beer et al., 2006, 2010; Hughes and Beer, in press; Somerville et al., 2010), but very little is understood about the psychological mechanisms that account for this association. An implication arising from previous research was that MOFC might mediate decision threshold shifts underlying the expression of self-serving cognition. The present study provided a direct test of this hypothesis by drawing on established behavioral methods that combine signal detection measurement of decision thresholds and a manipulation known to shift decision thresholds underlying self-serving responses. The findings (a) replicate previous work showing that increased activation in MOFC, LOFC, MPFC, and dACC was associated with conditions that attenuated self-serving evaluations (Beer and Hughes, 2010; Beer et al., 2010; Blackwood et al., 2003; Hughes and Beer, in press; Krusemark et al., 2008), and (b) build on previous work by showing that MOFC activation predicted individual differences in decision threshold shifts across conditions. Participants recruited more MOFC activation to the extent that they shifted their decision threshold in a less self-serving manner (i.e., more
conservative) when held accountable. The present findings contribute to the growing body of neural research on social cognition and have implications for conceptualizing the role of frontal lobe function in motivation and cognition interactions.

An integration of the current findings with the broader role of MOFC in contextual updating suggests a revised hypothesis regarding the impact of MOFC activation on decision threshold shifts underlying the expression of self-serving cognition. MOFC activation should not necessarily predict conservative decision threshold shifts or necessarily predict attenuation of self-serving responses. Instead, MOFC engagement should relate to decision threshold shifts and changes in the expression of self-serving tendencies as a function of updating the expression of automatic self-processes across contexts. For example, in contrast to the robust effect of increased MOFC engagement on conservative decision threshold shifts and attenuation of self-serving expression, a different pattern of results would be expected in conditions that motivate people to upregulate the expression of self-serving tendencies (e.g., self-esteem threat) or for populations that have negative automatic associations with self (e.g., low self-esteem or depressed individuals).

Research has demonstrated a broad role of MOFC in updating, that is, shifting the threshold at which prepotent tendencies are expressed as contexts change (Beer et al., 2004; Lhermitte, 1986; Stuss and Benson, 1984). Medial OFC activation has been associated with both the downregulation and the upregulation of behavior in relation to contextual changes (e.g., Bhanji and Beer, in press; Cooney et al., 2010; Hare et al., 2011; Mehta and Beer, 2009). Whereas both the current study and previous research on the general updating function of OFC find activation peaks in BA 11, it is important to keep in mind that the activation found in the current study is more ventral and anterior than some of the activation associated with the updating function. The activation in the current study is most similar to the region of OFC where patients tend to sustain brain damage either from traumatic brain injury, tumor resection or stroke. In fact, damage to the OFC, including the medial OFC, is also classically associated with the inflexible expression of prepotent or automatic tendencies even when they become contextually inappropriate (Lhermitte, 1986; Stuss and Benson, 1984 and see Beer et al., 2006; Fellows and Farah, 2003; Sellitto et al., 2011). Therefore, one way MOFC may affect self-judgments is by updating their underlying prepotent (or automatic) strategies or components.

When considered in the context of the current study, MOFC activation may have predicted decision threshold shifts not necessarily because they were conservative in nature but because they departed from prepotent thresholds for expressing baseline self-serving associations. There is strong evidence that self-serving cognition is prepotent or at least relies on relatively automatized processing. Most people have implicit and explicit positive associations with self (Gray-Little et al., 1997; Koole et al., 2001; Twenge and Campbell, 2008) and self-serving responses (more positive than warranted by external criteria) are pronounced or unaffected when people are placed under mental load (Aliche et al., 1995; Beer and Hughes, 2010; Koole et al., 2001; Kruger, 1999; Lench and Ditto, 2008; Paulhus et al., 1989). The pervasive baseline positivity of self-judgment may be why previous research and the current study tend to find that increased MOFC engagement predicts conservative shifts (i.e., stricter threshold for expressing baseline positive associations) that attenuate self-serving judgment.

For example, in the paradigm posed by the current study, people’s inclination is to claim as much knowledge as possible because it casts the self in a flattering light and, therefore, use liberal thresholds for their familiarity judgments (Paulhus et al., 2003). However, accountability introduces a new context and decision threshold may be updated (that is, shifted) to balance the baseline tendency to claim as much knowledge as possible against the possibility of making the self look foolish if weakly held knowledge is a mistake. In this way, MOFC activation may have marked shifts in decision threshold that were conservative (rather than liberal) because accountability made participants more conservative about acting on their baseline tendency to try to claim as much knowledge as possible.

However, if the association between MOFC engagement and self-serving responses reflects a departure from baseline strategies or components of self-judgment, then MOFC should predict both conservative and liberal shifts in decision threshold as long as they depart from baseline tendencies. We are not hypothesizing that MOFC activation simply predicts any change across an experimental manipulation. Instead, the recruitment of MOFC should be related to a predictable direction of threshold shift and change in self-serving responses as a function of relevant automatic self-processing. The accountability and overclaiming paradigm used in the current study is not suitable for testing this hypothesis because it primarily elicits one combination of decision threshold shift and change in self-serving responses (and see Paulhus et al., 2003; Sedikides et al., 2002). However, previous behavioral research provides clear avenues for testing this hypothesis in future neural research by identifying situations and populations that decouple the direction of decision threshold shift and direction of its impact on the expression of self-serving responses.

For example, people with moderate to high self-esteem respond to self-esteem threats by upregulating their baseline self-serving tendencies. When they fail at a task or are rejected by peers, people show an even greater tendency to exaggerate their own desirability when comparing their personality to peers (Hughes and Beer, submitted for publication; Vohs and Heatherton, 2004). If MOFC activation mediates threshold shifts away from baseline, then it should be associated with a shift towards a more liberal threshold for expressing positive information about the self in the face of threat, and subsequently, increased self-serving responses (Hughes and Beer, submitted for publication). Conversely, depressed people or people with low self-esteem tend to have baseline associations with self that are negative rather than positive (e.g., Koole et al., 2001 and see Phillips et al., 2010 for a review) and can respond to self-esteem threats with increased self-deprecation (e.g., Vohs and Heatherton, 2004 and see vanDellen et al., 2011 for a review). In these populations, MOFC engagement should predict the combination of a shift towards a more liberal threshold underlying an increase in responses that are self-deprecating (rather than self-serving) as a function of self-esteem threat. If these hypotheses were borne out, those results would suggest that the one reason for the association between MOFC and self-serving responses is its role in updating

Table 2

<table>
<thead>
<tr>
<th>Region of activation</th>
<th>BA</th>
<th>Coordinates (x, y, z)</th>
<th>t-Stat</th>
<th>Cluster size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medial orbitofrontal (L)</td>
<td>11</td>
<td>−6 58 −20 6.62</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>Medial orbitofrontal (R)</td>
<td>47</td>
<td>32 44 −18 5.75</td>
<td>208</td>
<td></td>
</tr>
<tr>
<td>Lateral orbitofrontal (L)</td>
<td>47</td>
<td>−30 56 −12 6.96</td>
<td>717</td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal (L)</td>
<td>10</td>
<td>−16 64 −2 6.12</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>Medial prefrontal (L)</td>
<td>8/9</td>
<td>−2 34 62 4.95</td>
<td>235</td>
<td></td>
</tr>
<tr>
<td>Dorsal anterior cingulate (L)</td>
<td>24</td>
<td>−4 14 34 5.44</td>
<td>174</td>
<td></td>
</tr>
</tbody>
</table>

BA = Brodmann’s area; all regions are significant at p < .05 corrected for multiple comparisons at the cluster-level.
baseline decision thresholds used for self-judgments rather than shifting thresholds in a necessarily conservative direction.

The findings from the present study may also contribute to ongoing discussions about the differential roles of MOFC and MPFC in self-processing (e.g., Beer, 2007; D’Argembeau et al., 2012; Stuss and Benson, 1984). Whereas previous research suggests a role for MPFC in accessing or representing internal information about the self (compared to other kinds of information) relevant to an evaluation about the self or other people (Beer et al., 2010; Jenkins et al., 2008), the present study suggests that MOFC serves to adjust the threshold for which that information is expressed in judgment as a function of context. This perspective may explain why dorsal MPFC activation differentiates self-judgments of personality that are more certain from those that are more uncertain while MOFC activation differentiates judgments about traits that are deemed important to possess compared to traits deemed unimportant to possess (D’Argembeau et al., 2012). MPFC modulation of certainty about self-judgment may reflect a relation between certainty and the extent to which relevant introspective information is accessible or represented. In contrast, MOFC activation may reflect decision threshold shifts that are modulated by trait importance. Behavioral research shows that people use more liberal definitions when judging traits they wished they possessed and these liberal definitions are associated with increased judgments of trait self-descriptiveness (e.g., Dunning, 1995; Suls, 1999). However, whereas both the current study and the study on trait importance report peaks in BA 11, the activation in the current study spreads in a more ventral direction. Therefore, more research is needed to fully understand whether MOFC is a region that modulates the expression of self-representations mediated by MPFC.

Table 3

<table>
<thead>
<tr>
<th>Region of activation</th>
<th>BA</th>
<th>Coordinates</th>
<th>t-Stat</th>
<th>Cluster size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occipital (R)</td>
<td>18/19</td>
<td>26 -96 14</td>
<td>8.98</td>
<td>1294</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32 -80 -6</td>
<td>6.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24 -84 -6</td>
<td>5.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-28 -72 -10</td>
<td>6.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-20 -78 -8</td>
<td>6.77</td>
<td></td>
</tr>
<tr>
<td>Occipital (L)</td>
<td>18</td>
<td>-18 -92 -8</td>
<td>7.09</td>
<td>1324</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-28 -72 -10</td>
<td>6.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-20 -78 -8</td>
<td>6.77</td>
<td></td>
</tr>
<tr>
<td>Lateral prefrontal (R)</td>
<td>45</td>
<td>50 28 2</td>
<td>6.27</td>
<td>515</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48 20 0</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58 24 10</td>
<td>5.41</td>
<td></td>
</tr>
<tr>
<td>Inferior parietal (L)</td>
<td>40</td>
<td>-58 -42 50</td>
<td>5.56</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-56 -50 42</td>
<td>3.45</td>
<td></td>
</tr>
</tbody>
</table>

BA = Brodmann’s area; Cues did not significantly account for the activation seen in the experimental trials; no significant clusters were found in the planned ROI analyses. To illustrate that the catch blocks did elicit differences, whole brain analyses were conducted; all regions are significant at p<.05 corrected for multiple comparisons at the cluster-level.

Fig. 3. Neural activation defined by the Accountable > Unaccountable contrast; parameter estimates plotted in relation to baseline for each condition. (*) indicates p<.05. BA = Brodmann’s area; MOFC = medial orbitofrontal cortex; LOFC = left lateral orbitofrontal cortex; ROFC = right lateral orbitofrontal cortex; MPFC = medial prefrontal cortex; dACC = dorsal anterior cingulate cortex.
Fig. 4. Individual differences in conservative shifts in decision threshold (c) (i.e., less overclaiming) modulate MOFC activation. Scatterplot depicts individual differences in decision threshold shifts as a function of Accountability ( Accountable decision threshold − Unaccountable decision threshold) in relation to parameter estimates from the MOFC activation identified in the Accountable > Unaccountable contrast.

Finally, another important direction for understanding the neural underpinnings of self-serving cognition is to understand why lateral OFC, MPFC, and dACC tend to show main effects of increased activation in various conditions that attenuate self-serving cognition but do not significantly predict individual differences in attenuation (e.g., Beer and Hughes, 2010; Beer et al., 2010; Hughes and Beer, in press). Research in other domains has associated many of these neural regions with difficulty or conflict (e.g., Beer et al., 2004; Botvinick et al., 2004; Grinband et al., 2011) yet that account does not readily fit with the existing research on self-serving cognition. For example, if conditions that attenuate self-serving cognition were simply more difficult or required more effort to resolve some kind of conflict, it would be reasonable to expect that these conditions would be associated with longer reaction times. However, that is not the case in the current study or in previous research (e.g., Beer et al., 2010; Hughes and Beer, in press). In the current study, difficulty might also be indexed by how difficult it was for participants to discriminate between existent and nonexistent items (‘d’). Yet activation found in the lateral OFC, MPFC, and dACC did not show significant correlations with the discriminability measure. Future research is needed to more directly target the function of these regions in relation to conditions that attenuate self-serving cognition.

Conclusions

In conclusion, the present study deepens our understanding of the processes supported by neural regions in self-serving cognition. MOFC activation is associated with changes in self-serving responses in a number of domains (Beer and Hughes, 2010; Beer et al., 2006, 2010; Blackwood et al., 2003; Hughes and Beer, in press; Somerville et al., 2010) and the present study demonstrated this relation in a new domain: self-serving claims of scholarly knowledge. Furthermore, the present study shows that one way MOFC recruitment attenuates self-serving responses is to shift decision thresholds in a more conservative manner. The findings open up new avenues for conceptualizing the role of MOFC in supporting contextually appropriate self-evaluations and behavioral selection. For example, in the present research, the MOFC activation is theorized to mark a ‘set shift’ in decision threshold with a signal that is sustained within a context (rather than transient signal elicited in a trial-by-trial manner). However, the blocked design in the current study makes it impossible to tease apart activation that represents tonic activation for each block from activation that reflects trial-by-trial changes. Therefore, future research using a mixed blocked, event-related design is needed to test these two alternate accounts (Visscher et al., 2003). Finally, these findings also contribute to developing neural models of social cognition which will benefit from considering how the neural regions identified in studies of non-motivated self-processing interact (or do not) with neural regions identified in studies of motivated self-processing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2012.03.011.

References


