Socially excluded individuals fail to recruit medial prefrontal cortex for negative social scenes

Katherine E. Powers, Dylan D. Wagner, Catherine J. Norris, and Todd F. Heatherton
Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

Converging behavioral evidence suggests that people respond to experiences of social exclusion with both defensive and affiliative strategies, allowing them to avoid further distress while also encouraging re-establishment of positive social connections. However, there are unresolved questions regarding the cognitive mechanisms underlying people’s responses to social exclusion. Here, we sought to gain insight into these behavioral tendencies by using functional magnetic resonance imaging (fMRI) to examine the impact of social exclusion on neural responses to visual scenes that varied on dimensions of sociality and emotional valence. Compared to socially included participants, socially excluded participants failed to recruit dorsomedial prefrontal cortex (dmPFC), a brain region involved in mentalizing, for negative social scenes. Moreover, following social exclusion, dmPFC demonstrated a linear effect of valence, with greater activity to positive social scenes compared to negative social scenes. These results suggest that, following social exclusion, people display a preference for mentalizing about positive social information and tend to avoid negative aspects of their social world.

Keywords: social exclusion; medial prefrontal cortex; withdrawal; mentalizing; social cognition

INTRODUCTION

As a social species, humans have a fundamental need to belong. Baumeister and Leary (1995) proposed that developing and maintaining interpersonal relationships evolved as an adaptive mechanism to promote survival. Feeling socially excluded, which threatens the need to belong, has been associated with self-defeating behaviors, negative moods, and mental and physical health problems (Twenge et al., 2001; Cacioppo et al., 2006). In order to avoid these adverse consequences, it is critical that humans are able to efficiently recognize and respond to threats of social exclusion. One theoretical model (Mitchell and Heatherton, 2009; Heatherton, 2011) suggests that the ability to detect and respond to exclusionary social threats relies on the ability to infer the thoughts, feelings and beliefs of others, a capability often referred to as having theory of mind or mentalizing (Gallagher and Frith, 2003; Mitchell et al., 2006; Gobbini et al., 2007). Critically, theory of mind allows people to recognize when evaluative judgments are possible and thus detect threats to inclusionary status. Due to the need to belong, these threats may motivate people to regulate their subsequent behaviors in order to forestall rejection or maintain their social bonds (Leary et al., 1995; MacDonald and Leary, 2005). However, there are inconsistencies in the extant research investigating responses to social exclusion.

Some prior research suggests that social exclusion motivates withdrawal from the surrounding world and leads to feelings of emotional detachment or ‘numbness’ (Twenge et al., 2001; Baumeister et al., 2002). Consistent with this reasoning, socially excluded individuals exhibit reduced empathic concern for the plight of others (DeWall and Baumeister, 2006) and are less likely to engage in prosocial helping behavior (Twenge et al., 2007). In other words, social exclusion apparently leads people to avoid others and to have less concern for them.

In contrast to the above research, other studies have found that social exclusion motivates participants to engage in affiliative behaviors and seek out positive social interactions. Gardner and colleagues demonstrated that social exclusion increases attention to social stimuli, as evidenced by enhanced memory for social information and events, as well as improved accuracy identifying facial expressions and emotional tones of spoken words (Gardner et al., 2000; Picket et al., 2004). Similarly, Maner et al. (2007) showed that excluded participants displayed a greater desire to make new friends and form positive impressions of others.

More recently, researchers have suggested that people might employ concurrent strategies to protect the self from further distress, while simultaneously attempting to form positive social connections with others (Hess and Pickett, 2010). From this perspective, people avoid potentially negative social interactions while paying greater attention to those that suggest more positive social outcomes. In line with this reasoning, recent research by DeWall et al. (2009) using an eye-tracking paradigm found that excluded
participants displayed decreased attention to negative social stimuli while selectively attending to signs of social acceptance.

These findings indicate that people’s responses to social exclusion are more nuanced than simply avoiding or approaching others. However, there are unresolved questions regarding the cognitive processes that underlie individuals’ differential reactions to social exclusion. As such, exploring neural mechanisms has the potential to provide insight into the nature of these cognitive responses. For instance, after experiencing an exclusionary social threat people may differentially engage attributional processes, or mentalize, to understand the intentions of others in their surrounding social world. The emerging neuroimaging literature on theory of mind has consistently implicated the medial prefrontal cortex as a central component of the neural systems that support such mentalizing (Frith and Frith, 2001; Mitchell et al., 2002; Gallagher and Frith, 2003; Mitchell et al., 2006; Gobbini et al., 2007). Although the medial prefrontal cortex is large and has been implicated in multiple distinct tasks, neuroimaging findings have converged on a specific dorsal region of the medial prefrontal cortex (dmPFC) as being critically involved in mentalizing about social knowledge (for reviews, see Mitchell, 2008; Lieberman, 2010). Recently, we demonstrated that people spontaneously recruit this particular region of dmPFC when viewing natural social scenes (Wagner et al., 2011), suggesting that complex social information generally promotes mental state attribution. Accordingly, dmPFC may index the extent to which people engage in mental state attribution for different types of social information and this may vary as a function of social relation status.

Given the behavioral research reviewed above, we propose that, following social exclusion, people should be more motivated to engage in mental state attribution for positive compared to negative social information. The goal of the present study was to test this hypothesis regarding the consequences of social exclusion by using functional magnetic resonance imaging (fMRI) to examine how neural responses to social cues are affected by social exclusion. If social exclusion promotes a subsequent desire to renew social connections, one might expect greater activity in the region of dmPFC associated with mentalizing when viewing positive social scenes. Furthermore, if social exclusion motivates withdrawal from signs of further social threat, one would expect less recruitment of this region of dmPFC when viewing negative social scenes. Here, we investigate these predictions by comparing neural responses of a group of participants who received a threat of future social exclusion to another group who believed they were likely to be socially included in the future.

MATERIALS AND METHODS

Participants

Thirty-four (22 females, age range 18–21 years) Dartmouth College undergraduates participated in this study. All participants were right-handed, had no history of neurological problems and had normal or corrected-to-normal vision. They received course credit or were paid for their participation and gave informed consent in accordance with the guidelines set by the Committee for the Protection of Human Subjects at Dartmouth College.

Materials and procedure

In order to experimentally manipulate social exclusion, we employed a modified version of the future alone versus future belonging paradigm (Twenge et al., 2001) adapted for the MRI environment. This paradigm has been used extensively in past research to induce feelings of social exclusion and isolation (Baumeister et al., 2002; DeWall & Baumeister, 2006; Maner et al., 2007; Twenge et al., 2001, 2002, 2003, 2007; DeWall et al., 2008, 2009).

When participants arrived, they were informed that this study was examining the effect of personality on ecological perception. Once in the scanner, participants completed a computerized version of the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975) and then were provided with feedback they believed was derived from their answers. In reality, feedback was randomly assigned prior to the experimental session. Half of the participants (n = 17) were told their future lives would be isolated and lonely (social exclusion), while the other half (n = 17) were told that theirs would be filled with long-lasting, stable relationships (social inclusion). To increase believability, the feedback also included personality descriptions typically believed by the average person (i.e. the ‘Barnum Effect’; Snyder, Shenkel and Lowery, 1977), as well as two statements regarding self-esteem and social skills accurately tailored to each participant (based on self-report information obtained in a mass survey). Immediately following feedback, participants completed a 24-item mood questionnaire (Vohs and Heatherton, 2001) to assess effectiveness of the manipulation.

Participants then underwent functional magnetic resonance imaging while viewing pictures selected from the International Affective Picture System (IAPS; Lang et al., 2008) and categorizing each as an indoor or outdoor scene (a task chosen to bolster the cover story and minimize the likelihood that participants would infer the true purpose of the study). Pictures varied on dimensions of sociality (social and non-social) and valence (negative, neutral and positive) and were matched for arousal and basic visual properties (i.e. luminance, hue, saturation and RGB values). Example stimuli include: social negative (a funeral; a domestic dispute), social neutral (buying groceries; talking on the phone), social positive (children playing at a water park; a romantic dinner), non-social negative (a burning building; a car accident), non-social neutral (a stack of books; a spoon) and non-social positive (a beautiful landscape; a delicious dessert). Critically, the non-social pictures did not contain any people. A total of 150 pictures (25 per category) were...
presented for 2.5s each. The order of the pictures was pseudo-randomized and counterbalanced across participants. In order to accurately estimate the hemodynamic response function, pictures were intermixed with passive fixation trials of variable durations (0–7500 ms). To minimize interruptions following the social exclusion manipulation, all pictures were presented in one functional run.

**fMRI procedure and analysis**

Structural and functional data were collected on a Phillips Intera Achieva 3T scanner at Dartmouth College using an eight-channel phase arrayed coil. An Epson ELP-7000 LCD projector was used to project stimuli onto a screen at the end of the magnet bore that participants viewed via an angled mirror mounted on the head coil.

Structural images were acquired using a T1-weighted MP-RAGE protocol (160 sagittal slices, TR = 9.9 ms, TE = 4.6 ms, 8° flip angle, 1 x 1 x 1 mm voxels). Functional images were acquired using a T2*-weighted echo-planar sequence (TR = 2500 ms, TE = 35 ms, 90° flip angle and field of view = 24 cm). Data were collected in one functional run consisting of 270 whole-brain volumes (36 axial slices per volume, 3 mm thick, 0.5 mm gap, 3 x 3 mm in-plane resolution).

Neuroimaging data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK). First, functional data were preprocessed to remove sources of noise and artifact and corrected for differences in slice acquisition time. Images were then realigned within the functional run to correct for head movement and unwarp to reduce residual movement-related image distortions not corrected by realignment. Functional data were then normalized into standard space (3 mm isotropic voxels) based on the SPM8 EPI template that conforms to the ICBM 152 brain template (Montreal Neurological Institute, MNI). Finally, normalized data were spatially smoothed using a 6 mm full-width-at-half-maximum Gaussian kernel.

A general linear model (GLM) incorporating task effects and covariates of non-interest (linear trend, six movement parameters derived from realignment) was specified for each participant. Each GLM was convolved with a canonical hemodynamic response function (HRF) and used to generate a contrast image comparing social to non-social activations for each participant. These contrast images were collapsed across groups and entered into a second-level random effects analysis, thresholded at $P < .001$ with an extent threshold of 10 contiguous activated voxels. This analysis resulted in a whole-brain statistical parametric map identifying regions displaying greater activity to social than non-social scenes.

In order to investigate the between-group differences of social exclusion vs social inclusion, we performed a region-of-interest (ROI) analysis on regions identified by this contrast. Parameter estimates ($\beta$) for each participant were extracted by centering a 6 mm sphere on the voxels of peak activation of all regions. ROIs were thus defined in an unbiased manner, as both groups (socially excluded and socially included) contributed equally to the statistical parametric map used for ROI identification. Parameter estimates were submitted to offline statistical analyses to examine effects of group and stimulus valence on regional brain responses to social vs non-social scenes.

**RESULTS**

**Behavioral results**

Analysis of the 24-item mood questionnaire revealed that socially excluded participants reported being in a more negative mood than socially included participants, $t(32) = 2.86$, $P = 0.007$; Mean$\text{excluded} = 64.3$, s.d. $= 12.2$; Mean$\text{included} = 75.3$, s.d. $= 10.0$.

**fMRI results**

A whole-brain analysis comparing regions that displayed a greater response to social scenes compared to non-social scenes for all participants revealed a system of regions including the medial prefrontal cortex, amygdala, posterior cingulate cortex/precuneus and regions of the inferotemporal cortex (Figure 1 and Table 1). These are areas that are typically activated in response to viewing social scenes (e.g. Iacoboni et al. 2004; Wagner et al., 2011).

To investigate the effect of social exclusion on the neural response to social vs non-social scenes, a difference score for neural activity during social scenes relative to non-social scenes was calculated for each participant for all valence categories (negative, neutral and positive) in each ROI identified in the previously described contrast. Positive difference scores indicate greater activity to social scenes, while negative difference scores indicate less activity to social scenes, both relative to non-social scenes of the same valence. For each ROI, these difference scores were interrogated using a mixed model ANOVA, with group (social inclusion and social exclusion) as a between-subjects factor and scene valence (negative, neutral and positive) as a within-subjects factor.

This analysis revealed that dmPFC was significantly modulated by social exclusion. Specifically, there was a main effect of valence, such that the difference in dmPFC activity between social and non-social scenes was greater for negatively valenced scenes than either neutral or positive scenes, $F(2,64) = 8.06$, $P = 0.001$. Moreover, there was a main effect of group, such that the difference between social and non-social scenes was greater for the socially included compared to socially excluded participants, $F(2,31) = 4.61$, $P = 0.039$. Importantly, these effects were qualified by a significant valence by group interaction, $F(2,64) = 3.82$, $P = 0.027$. For the negatively valenced pictures, socially excluded participants demonstrated less dmPFC activity to social than non-social scenes (negative difference scores), whereas socially included participants showed greater dmPFC activity to social than non-social scenes.
scenes (positive difference scores). There were no between-group differences for positive and neutral scenes (both $P > 0.17$) (Figure 2).

Importantly, this effect was driven primarily by the social scenes. When the analysis was restricted to non-social scenes only, there were no between-group differences in negative, neutral or positive scenes in dmPFC (all $P > 0.40$). That is, the interaction pattern observed cannot be attributed to the non-social scenes.

We also performed a linear trend analysis on the neural difference scores in dmPFC to test the linear relationship across valence categories within each group. This analysis revealed a significant linear relationship ($P < 0.001$) for socially excluded participants. Specifically, as the valence of the scenes becomes more positive, dmPFC activity to social scenes increases. This relationship was not observed for socially included participants ($P = 0.40$) (Figure 2).

In order to investigate whether observed differences in dmPFC activity were attributable to differences in mood, we ran a regression analysis testing this meditational model. Consistent with prior behavioral research (e.g. Twenge et al., 2007; DeWall et al., 2009), results revealed that mood did not significantly mediate the relationship between social exclusion and dmPFC activity ($P = 0.78$).

Only one other region (orbitofrontal cortex, OFC) displayed a main effect of group, $F(1,32) = 4.69$, $P = 0.038$, such that socially included participants displayed greater activity to social than non-social scenes compared to the socially excluded participants. There was no main effect of valence ($P = 0.19$) or valence by group interaction in OFC ($P = 0.58$). No other regions showed significant main effects of group (all $P > 0.10$) or significant group by valence interactions (all $P > 0.15$).

**DISCUSSION**

Given the fundamental drive to establish and maintain social relationships, responding to exclusionary social threats is critical for our survival as a social species. Our results suggest that the threat of social exclusion alters processing of social cues such that people subsequently engage in more apparent mentalizing when viewing positive compared to negative social scenes. We propose that these strategies reflect differential engagement of neural regions involved in understanding and empathizing with others.
In line with prior research (e.g., Wagner et al., 2011), viewing social compared to non-social scenes activated the ‘social brain’, a system of brain regions that have consistently been shown to respond to social information (for reviews, see Lieberman, 2007; Adolphs, 2009; Heatherton, 2011). As predicted, a region of dmPFC was uniquely sensitive to social exclusion and this varied as a function of stimulus valence. Importantly, this specific region of dmPFC has been found across numerous studies to be implicated in social cognition and mentalizing (Mitchell, 2008; Lieberman, 2010) and overlaps with the area of dmPFC found in our prior work on spontaneous mentalizing (Wagner et al., 2011). Crucially, this effect was driven primarily by social scenes, as there was no effect of social exclusion on neural responses to non-social scenes.

Across numerous neuroimaging studies, this particular region of dmPFC has been implicated in thinking about the mental and emotional states of other people, both in explicit mentalizing tasks (Mitchell et al., 2002, 2004, 2006; for review, see Gallagher and Frith, 2003) and also when no mentalizing instructions are given (Iacoboni et al., 2004; Spiers and Maguire, 2006; Gobbini et al., 2007; Wagner et al., 2011). In keeping with this functional role of dmPFC and with previous behavioral research documenting the potential protective nature of withdrawal responses (e.g., Baumeister et al., 2002), we propose that the differential activation of dmPFC to negative social scenes following social exclusion is an adaptive response that serves to blunt negative aspects of the social world. This interpretation is consistent with an eye-tracking study by DeWall and colleagues (2009) demonstrating that socially excluded participants attended more to positive than negative social information.

However, our findings are indicative of more than merely attentional avoidance. Indeed, medial prefrontal regions implicated in attention and executive control processes are consistently more dorsal than the region reported here (Wager et al., 2004). We speculate that the failure to recruit this particular region of dmPFC following social exclusion suggests a desire to avoid considering the mental states of potential social threats. More broadly, this pattern of activity may reflect a general desire to avoid any additional socially derived negative emotion after receiving an exclusionary social threat. Likewise, as predicted by behavioral research, socially excluded participants displayed increased dmPFC activity to visual depictions of positive social scenes, suggesting that social exclusion leads to greater mentalizing for positive social cues.

The specific region of dmPFC reported here has been linked to empathic emotional inferences (Hooker et al., 2008; Rameson et al., 2012) and empathic responses to seeing others experiencing rejection (Masten et al., 2011). Moreover, previous research has demonstrated that activity within this particular region is strongly correlated with individual differences in trait empathizing (Wagner et al., 2011) and has been implicated in making accurate empathic inferences for the emotions of others (Zaki et al., 2009). Additionally, DeWall and Baumeister (2006) found that socially excluded people reported less empathic concern for the social misfortunes of others, such as being rejected by a romantic partner. Taken together, these findings lend credence to our conjecture that reduced dmPFC activity following social exclusion reflects less mentalizing for the contents of negative social scenes.

By providing evidence of a nuanced neural pattern of responses to social exclusion, our results extend existing
theories hypothesizing both approach and avoidance responses to social exclusion (e.g. Hess and Pickett, 2010; Maner et al., 2007) by demonstrating that these strategies reflect differential engagement of neural regions involved in understanding and empathizing with others. Moreover, our results are consistent with a theoretical model proposing that both detecting and responding to social threat modulates activity in dmPFC (Heatherton, 2011). These results suggest that threats to social exclusion instigate regulatory responses to cope with potential threat. One of the earliest models of self-regulation (Carver and Scheier, 1982) proposed that people continue to regulate their behaviors in adaptive and profitable ways when favorable outcomes are expected. However, when unfavorable outcomes are expected, people escape from self-awareness and withdrawal from further attempts. Our findings converge nicely with the Carver and Scheier model, since unfavorable outcomes following social exclusion (e.g. thinking about the mental states of potential social threats) are met with mental withdrawal, while favorable outcomes (e.g. considering re-establishing social connections) are met with continued, possibly enhanced efforts.

Our findings provide insight into the cognitive processes underlying behavioral consequences of social exclusion. We speculate that social exclusion results in a suspension of the attributional processes that are typically deployed when attempting to understand social situations, at least for negative social information. In this way, failure to mentalize about negative aspects of the social world may undermine future behavioral efforts. Such an interpretation is consistent with behavioral studies that have shown that social exclusion leads to self-defeating behaviors such as aggression (Twenge et al., 2001) and decreased empathy for others (DeWall and Baumeister, 2006).

Given the lack of additional control conditions, it can be difficult to separate the effects of social exclusion from changes in overall mood. However, we note that prior research using this paradigm has consistently failed to find any effects of generalized negative mood (see Twenge et al., 2001, 2002, 2003, 2007; Baumeister et al., 2002; DeWall and Baumeister, 2006; Maner et al., 2007; DeWall et al., 2009). Thus, receiving negative feedback without an exclusionary social threat component has not been shown to significantly affect participants’ behavior or cognitions. Furthermore, in the present study mood failed to mediate the relationship between social exclusion and dmPFC activity, providing further evidence that mood is not driving the observed effects.

Likewise, the lack of a neutral feedback condition may raise the possibility that the observed effects result from feeling socially included, rather than excluded. Again, we note that previous studies with neutral feedback conditions repeatedly find no differences between this condition and the social inclusion condition (Baumeister et al., 2005; DeWall and Baumeister, 2006), suggesting that this condition functions as a reliable control. Indeed, previous work has directly compared exclusionary to inclusionary social feedback in the absence of any additional conditions (Twenge et al., 2007). Thus, in light of the consistency demonstrated by previous research using this paradigm, no other feedback conditions were included in the present study.

In summary, we used fMRI to examine the neural bases underlying the notion that social exclusion results in both defensive and affiliative responses (Hess and Pickett, 2010). Our results confirm previous research suggesting that people are motivated both to protect themselves from further distress (Twenge et al., 2001; Baumeister et al., 2002; DeWall and Baumeister, 2006) and seek out positive social connections (Gardner et al., 2000; Picket et al., 2004; Maner et al., 2007) following exclusionary social threats. Moreover, our results suggest that these responses to social exclusion may be explained as a reduced motivation to mentalize about negative aspects of the social world.

REFERENCES


Neural responses to social exclusion


